# Psychological Review

THEODORE M. NEWCOMB, Editor
*University of Michigan*

## CONTENTS

---

# THE PSYCHOLOGICAL REVIEW

## SYMMETRIC UNCERTAINTY ANALYSIS AND ITS IMPLICATIONS FOR PSYCHOLOGY [1]

### W. R. GARNER

*Johns Hopkins University*

Information theory, as developed primarily by Shannon (**8**), has had considerable impact on psychological research. Information theory has been used as a model for many different types of behavior; it has also provided psychologists with a statistical measure which has many useful properties. Frequently the measure is used where the term *information* is somewhat inappropriate, and for this reason Garner and McGill (**3**) have suggested that the term *uncertainty* be applied to the measure to divorce its statistical properties from the content implications of the term *information*. Even when the mathematical properties of the uncertainty measure lead to implications for behavior situations, information theory does not necessarily define the problem; rather,

the mathematical properties of the uncertainty measure simply make certain relations clear which other statistical techniques have not been able to do. The purpose of this paper is to develop some equations of this sort involved in uncertainty analysis, and to point out the relevance of these equations to a few psychological problems.

Two valuable properties of the measure of uncertainty, as pointed out by Garner and McGill (**3**), are: (*a*) that it is a nonparametric measure which can be applied to any set of categorized data, and (*b*) that the uncertainty of a variable can be partitioned into component parts, much as variance is partitioned in analysis of variance. A third useful property of the uncertainty measure is that an uncertainty analysis can be carried out in a completely symmetric form, in which it is unnecessary to distinguish between meanings or functions of the several variables involved in the analysis. This last property derives in part from the nonparametric nature of the measure, and is not a property of analysis of variance, in which different properties of the criterion and predictor variables are assumed.

## Analysis of Three Variables

For illustration purposes, let us consider first a three-dimensional matrix of data. Figure 1 shows schematically such a data array and some of the terms which we will be using. The three variables are $w$, $x$, and $y$, and they can assume values $w_i$, $x_j$, and $y_k$. Each cell in the matrix originally contains a number which represents the frequency of occurrence of cases having that particular combination of values of $w$, $x$, and $y$. For our purposes, we have shown the matrix in which these frequencies have been transformed into proportions, $p(i,j,k)$, by dividing the cell frequencies by the total number of cases in the matrix. This matrix can be collapsed across any one variable to form three different two-variable matrices, or it can be collapsed across two variables to give distributions for one variable at a time. When two-variable matrices
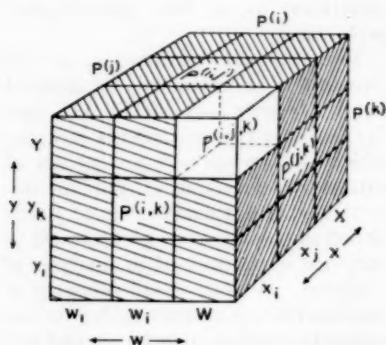


Fig. 1. Symbolic representation of a three-variable matrix of data. Three variables, $w$, $x$, $y$, have category values $w_i$, $x_j$, and $y_k$. Each cell in the matrix originally has a frequency which is divided by the total number of cases in the matrix to give the proportion $p(i, j, k)$. Summing these proportions across any variable gives two-variable matrices with cell entries of $p(i, k)$, $p(j, k)$, or $p(i, j)$ as indicated on the three surfaces. Summing across two variables gives the marginal proportions $p(i)$, $p(j)$, and $p(k)$.

are used, the proportions in the cells are still computed as the frequency in a cell divided by the total number of cases, and can be obtained by summing the proportions across the collapsed variable.

*Unidirectional analysis.* When uncertainty is partitioned as in analysis of variance (3, 4), we start by identifying one of the variables, let us say $y$, as the criterion variable and the other two, $w$ and $x$, as predictor variables. Now our problem is to partition the uncertainty of $y$ into its component parts. The basic partitioning equation is

$$U(y) = U(y:w,x) + U_{wx}(y) \quad [1]$$

where $U(y)$ is the uncertainty of the $y$-distribution, based on $p(k)$; $U(y: w,x)$ is the total uncertainty in $y$ which is predictable from $w$ and $x$, and it is called a multiple contingent uncertainty; and $U_{wx}(y)$ is the residual or error uncertainty.[3] This last term is a conditional uncertainty, and is computed by taking one combination of $w$ and $x$ at a time, determining the uncertainty of the $y$ distribution, and then obtaining a weighted average for all of the $w,x$ cells.

By transposing terms in Equation 1, we can define the multiple contingent uncertainty as

$$U(y:w,x) = U(y) - U_{wx}(y) \quad [2]$$

The right hand side of this equation can then be rewritten to show the complete partitioning of the predictable uncertainty as

$$U(y:w,x) = U(y:w) + U(y:x) + U(y:\overline{wx}) \quad [2a]$$

[3] For many of the equations presented here we shall not present calculational formulas or complete proofs, since these are available in at least a related form in other articles (especially 2, 3, 4, 5).

The three terms on the right of Equation 2a represent, in order, the uncertainty in $y$ which is predictable from $w$, that which is predictable from $x$, and that which is preditable from unique combinations of $w$ and $x$. The first two terms are simple contingent uncertainties, and are normally computed from collapsed two-variable matrices. The last term is an interaction uncertainty, and is computed either as a residual or as the difference between a simple contingent uncertainty and the same contingent uncertainty computed with the third variable held constant, rather than with the matrix collapsed over that variable. (See **3**, **4**.)

The partitioned uncertainties on the right of Equation 2a are completely analogous to partitioned variances in analysis of variance, and would in that case be called the main effect due to $w$, the main effect due to $x$, and the interaction between $w$ and $x$.

*Symmetric analysis.* This approach to uncertainty analysis has the advantage of being completely analogous to the more familiar analysis of variance. It is possible, however, to approach uncertainty analysis from another point of view which allows us to deal with the total matrix of data as a unit. We can call this type a symmetric uncertainty analysis to distinguish it from the unidirectional analysis in which the uncertainty of just one variable is partitioned.

Since uncertainty analysis is nonmetric, we can look at the three-dimensional matrix shown in Fig. 1 as a single frequency distribution in which we have categories and proportions of total cases falling in each category. Thus we can compute the uncertainty of the total matrix of data,

$$U(w,x,y) = - \Sigma\{[p(i,j,k)] \times [\log_2 p(i,j,k)]\} \quad [3]$$

where $U(w,x,y)$ is the total uncertainty in the three-dimensional matrix, and $p(i,j,k)$ is as defined above. Now this total uncertainty can be broken down in many different ways, some of them being

$U(w,x,y)$

$$= U(y) + U_y(x) + U_{xy}(w) \quad [4a]$$
$$= U(w) + U_w(x) + U_{wx}(y) \quad [4b]$$
$$= U(w,x) + U_{wx}(y) \quad [4c]$$
$$= U(w,y) + U_{wy}(x) \quad [4d]$$
$$= U(x,y) + U_{xy}(w) \quad [4e]$$

For these equations, the only terms needing additional explanation are $U_y(x)$ or equivalent forms, and $U(w,x)$ or equivalent forms. Any term of the form $U_y(x)$ is a conditional uncertainty, and is computed by taking one value at a time of the subscripted variable, determining the uncertainty of the variable in parentheses, then obtaining a weighted average over the subscripted variable. For such a calculation, the matrix is collapsed over the third variable. A term of the form $U(w,x)$ is a total uncertainty for a two-dimensional matrix, i.e., one collapsed over the third variable. It is computed from the $p(i,j)$, in this case, or the appropriate terms from any other combination of two variables.

$U(w,x,y)$ is a measure of the total uncertainty actually obtained in the matrix. We can also determine the maximum total uncertainty which could have been obtained from the matrix by determining what proportion of cases would fall in each cell if there were no contingencies between any of the variables. This ideal proportion, $P(i,j,k)$, is

$$P(i,j,k) = p(i)p(j)p(k) \quad [5]$$

in much the same way as is done for chi square. From these ideal propor-

tions, representing those which would be obtained in a matrix with all variables completely independent, we can compute the maximum uncertainty in the matrix, $U_{max}(w,x,y)$, as

$$U_{max}(w,x,y) = -\Sigma\{[p(i)p(j)p(k)] \times [\log_2 p(i)p(j)p(k)]\} \quad [6]$$

This maximum uncertainty can itself be partitioned into components, and in this case the partitioning results in a simple set of terms.

$$U_{max}(w,x,y) = U(w) + U(x) + U(y) \quad [7]$$

In other words, the maximum uncertainty which can be obtained in a matrix which is restricted only by the marginal totals is the sum of the uncertainties for the three variables taken one at a time.

Now if we have the maximum uncertainty which can be obtained in a matrix, and the uncertainty which actually is obtained, the difference between them must give us a measure of the total contingency or interrelatedness in the three-dimensional matrix. We can define a measure of this total contingency as

$$U(w:x:y) = U_{max}(w,x,y) - U(w,x,y) \quad [8]$$

The term on the left is the total contingent uncertainty in the matrix, and the colon notation has been used as for other contingent uncertainty terms. It should be noted, however, that there is no implication of direction in the notation; the three symbols for the three variables can be written in any order.

The total contingent uncertainty can be partitioned also, and the use of Equation 7 with Equations 4c, 4d, or 4e leads to three different forms of the partitioning.

$$U(w:x:y)$$
$$= U(w:x) + U(y:w,x) \quad [9a]$$
$$= U(w:y) + U(x:w,y) \quad [9b]$$
$$= U(x:y) + U(w:x,y) \quad [9c]$$

All terms in these equations have been defined above, and need no further clarification here. These equations show that the total contingent uncertainty can be partitioned into a multiple contingent uncertainty in which one variable is predicted from the other two, plus a simple contingent uncertainty between the two variables used as predictors, and it makes no difference which variable is treated as a criterion and which as predictors for this statement to be true.

As in Equation 2a, each of the three multiple contingent uncertainties can itself be partitioned into components, so that the total contingent uncertainty can be partitioned more completely.

$$U(w:x:y) = U(w:x) + U(w:y) + U(x:y) + U(y:\overline{wx}) \quad [10a]$$
$$= U(w:x) + U(w:y) + U(x:y) + U(x:\overline{wy}) \quad [10b]$$
$$= U(w:x) + U(w:y) + U(x:y) + U(w:\overline{xy}) \quad [10c]$$

Thus the total contingent uncertainty consists of the three simple contingent uncertainties involving the three pairs of predictors, plus the interaction term. Although the three interaction terms have been written differently for each of the above equations, it is clear from these equations that all three interaction uncertainties are equal. In other words, there exists in the matrix all of the two-variable contingencies, plus a single interaction term which truly represents a three-dimensional pattern of inconsistencies. For this reason, it would be better to write the interaction uncertainty as

$U(\overline{wxy})$ to make clear the uniqueness of that term for the given matrix.

We can transpose the terms in Equation 9 to see the relation of each of the three possible multiple contingent uncertainties to the total contingent uncertainty, and we obtain

$$U(y:w,x) = U(w:x:y) - U(w:x) \quad [11a]$$

$$U(x:w,y) = U(w:x:y) - U(w:y) \quad [11b]$$

$$U(w:x,y) = U(w:x:y) - U(x:y) \quad [11c]$$

With the equations written in this form, some interesting relations between variables become clear. Each of these equations shows the multiple contingent uncertainty for predicting one of the three variables from the other two. The form of the equations is such that the amount of prediction available for each variable is equal to a constant less a variable term, which is in each case the contingent uncertainty between the two predictor variables. In other words, the amount of predictable uncertainty available for any variable in the matrix is the total contingent uncertainty less the contingent uncertainty between the two variables from which the predicting is being done. Therefore, the amount of prediction available in any direction in the matrix is inversely related to the contingency between the two predictors. For a given matrix the maximum possible predictability will be obtained when the predictor variables are orthogonal, i.e., have no contingent uncertainty between them. This seems intuitively sensible, and the uncertainty analysis presented here makes it clear that for any set of data this relation must hold.

In order to compute the exact amount of a multiple contingent uncertainty involving one criterion and two predictor variables, it is necessary to carry out calculations which involve the three-dimensional matrix. For many sets of data, such a calculation would be extremely laborious, if not prohibitive. Equation 11 shows that it is possible to determine how much better prediction is for one variable than for another without actually carrying out a complete three-dimensional calculation. The amount by which one of the three multiple contingent uncertainties is greater than another is exactly the same as the difference in the contingent uncertainties between the two sets of predictor variables, and these uncertainties can be calculated with the three-dimensional matrix collapsed to two dimensions. Actually, the interaction term is the one which requires a three-dimensional calculation to determine all the components of the multiple contingent uncertainty (see Equation 2a), and the essential fact which makes it possible to determine exact differences between multiple contingent uncertainties without the complete calculation is that the interaction uncertainty, as mentioned above, is the same for all three criterion variables. It is truly a term which applies to the complete matrix of data, and is not uniquely determined for each variable used as a criterion, as it would appear to be from a unidirectional partitioning approach to uncertainty analysis.

## HIGHER-ORDER MATRICES

The entire discussion so far has been concerned with three-dimensional matrices. It is necessary to use at least three to show the nature of the principles, and it is easier to use just three to avoid undue complication. However, a brief discussion of at least the four-variable case will

illustrate some additional factors. We noted above that it is necessary to have two-dimensional uncertainty calculations in order to determine the differences between the various three-dimensional, multiple contingent uncertainties. In other words, calculations with as many variables as there are in the matrix are necessary to obtain absolute numbers for the multiple contingent uncertainties, but calculations with one less variable suffice to determine the differences between the multiple terms. A similar situation exists for the general case, as is shown by the following equations, in which $v$ is the fourth variable:

$$U(y:v,w,x) = U(v:w:x:y)$$
$$- U(v:w:x) \quad [12a]$$

$$U(x:v,w,y) = U(v:w:x:y)$$
$$- U(v:w:y) \quad [12b]$$

$$U(w:v,x,y) = U(v:w:x:y)$$
$$- U(v:x:y) \quad [12c]$$

$$U(v:w,x,y) = U(v:w:x:y)$$
$$- U(w:x:y) \quad [12d]$$

Each of the four multiple contingent uncertainties is the total contingent uncertainty in the four-variable matrix, $U(v:w:x:y)$, less the total contingent uncertainty between the three variables used as predictors, and exact differences between the multiples can be obtained with three-variable calculations. Similar conclusions concerning the variable whose prediction will be maximum can be made, since any variable which can be predicted from orthogonal predictor variables will have all of the contingent uncertainty in the matrix available for prediction of that variable.

Many situations exist, of course, where the data matrix has four or more variables, but where it is impractical or impossible to compute interaction terms involving even three variables.

We noted above that when only three variables are involved, exact differences between the multiples can be determined because there is only one interaction term for the entire matrix. When four or more variables are involved, however, the interaction terms involved in each multiple contingent uncertainty are not identical. For example, with four variables, each multiple contingency requires four interaction terms, three involving three terms, and one involving all four terms. For any pair of multiple contingent uncertainties, three of these four interaction terms are identical, but the fourth will normally be different. As an example, the complete set of terms involved in two of the four multiple contingent uncertainties are

$$U(y:v,w,x) = U(y:v) + U(y:w)$$
$$+ U(y:x) + U(\overline{vwy}) + U(\overline{vxy})$$
$$+ U(\overline{wxy}) + U(\overline{vwxy}) \quad [13a]$$

$$U(x:v,w,y) = U(x:v) + U(x:w)$$
$$+ U(x:y) + U(\overline{vwx}) + U(\overline{vxy})$$
$$+ U(\overline{wxy}) + U(\overline{vwxy}) \quad [13b]$$

where $U(\overline{vwxy})$ is the four-variable interaction. Notice that only the terms $U(\overline{vwx})$ and $U(\overline{vwy})$ are different interactions for these two multiples, and that the other three interactions in each case are identical.

This fact means that estimated differences between the various multiple contingent uncertainties can be obtained with little error simply by adding up the two-variable contingent uncertainties involved in each case. The total difference in the interaction uncertainties must be small, since only one of four terms is different. All four interaction terms, of course, are needed to determine the absolute magnitude of the multiple contingent uncertainty, and, therefore, more error in estimating the absolute magnitude

is likely in that case. When even more variables are included, the approximation by adding up the two-variable contingent uncertainties is more risky, since many more interaction terms are involved. However, it should be remembered (3, 4, 5) that interaction terms can take negative as well as postive values, so that it is often still reasonable to assume that the sum of all the interaction terms is negligible, even though any one of them may not be.

## PSYCHOLOGICAL APPLICATIONS

The mathematical developments briefly described in the preceding sections have many interesting applications and implications in psychological research. In some cases, their primary value is to clarify the meaning of computations which are commonly carried out with the uncertainty measure. In other cases, these equations lead to some quite specific predictions about behavior which is related to nonrandom sequential events. A few of these implications will be discussed as illustrative cases.

### Multivariate Information Transmission

One common application of information theory in psychology has been the use of information-transmission measures to determine the "channel capacity" of the human observer in transmitting, verbally or otherwise, information about a series of events or stimuli. Following the paradigm described by Garner and Hake (2), a typical experiment requires subjects to make an absolute response to a series of stimuli varying along some dimension, e.g., brightness or loudness, and the experimental result is expressed as an amount of information transmitted between stimuli and responses. If the information trans-

mitted is less than the information contained in the stimuli and in the responses, it is assumed to represent a channel capacity for the observer. In the more general terminology of the present paper, information transmission is a two-variable contingent uncertainty.

Although many experiments have been done using this approach, for illustrative purposes we will refer to one by Garner (1). In this experiment, observers were asked to make an identifying response on a twenty-point scale to twenty different intensities of a tone. From the two-dimensional matrix involving just stimuli and responses, the contingent uncertainty between stimuli and responses is computed, and this figure can be interpreted as information transmitted. For this two-variable case, it makes no difference whether we state that information has been transmitted from stimuli to reponses or vice versa, or whether we talk about the amount of uncertainty in stimuli which can be predicted from responses, or vice versa. The measure of information transmission is completely bilateral, and can be given interpretation in either direction.

Additional computations were made in this experiment, however, to allow multiple predictions. For example, the multiple contingent uncertainty for predicting the response from knowledge of the observer and of the stimulus was calculated, and this multiple contingent uncertainty was, naturally, larger than the simple contingent uncertainty. It seems quite reasonable that if more variables are isolated for prediction purposes, more prediction should be available.

However, Equation 11 shows that, once more than one variable is used as a predictor, the bilateral interpretation of the contingent uncertainty

is no longer possible. Let us look at this over-all situation as a three-variable matrix involving stimuli ($S$), responses ($R$), and observers ($O$). Now the multiple contingent uncertainty for predicting $R$ from $S$ and $O$ is the total contingent uncertainty in the matrix minus the contingent uncertainty between $S$ and $O$, the two predictor variables. In symbolic form,

$$U(R:S,O) = U(R:S:O) \\ - U(S:O) \quad [14a]$$

In this experiment, however, as in most such experiments, all observers received the same simuli. Therefore, there is orthogonality between $S$ and $O$, and the contingent uncertainty between them will be zero. Thus the contingent uncertainty available for predicting $R$ is the same as the total contingent uncertainty in the three-dimensional matrix, and is at a maximum for this particular matrix.

Suppose now that we decide to predict stimuli from responses and observers—the type of prediction which is more realistic, since in the ordinary situation the response is transmitted, and from it and knowledge of the observer we make inferences or predictions about the stimulus. In this case, in symbolic form,

$$U(S:R,O) = U(R:S:O) \\ - U(R:O) \quad [14b]$$

Again the multiple contingent uncertainty is the total contingent uncertainty in the matrix minus the contingent uncertainty between the two predictors, $R$ and $O$. But in this case there is no reason why $R$ and $O$ should be orthogonal, and, in fact, if they are, there is little point to adding $O$ as a predictor variable. Since $U(R:O)$ will be real and positive, prediction of $S$ from $R$ and $O$ must be less than pre-

diction of $R$ from $S$ and $O$, as in the former case.

These relations do not indicate that no improvement in prediction of $S$ can be obtained by making use of knowledge about $O$, but only that prediction of $S$ cannot be as great as prediction of $R$. If we consider only the prediction of $S$, then prediction from both $R$ and from $O$ can be broken down as follows:

$$U(S:R,O) = U(S:R) + U(S:O) \\ + U(\overline{SRO}) \quad [15]$$

The third term on the right is the interaction. This equation shows the problem clearly, since no prediction of $S$ can be obtained from just $O$ as long as $S$ and $O$ are orthogonal. Additional prediction may, however, be obtained from the interaction term, which is to say that only inconsistencies in relations involving observers are useful in predicting $S$. For predicting $R$, however, both inconsistent and systematic differences by $O$ in use of the $R$ continuum provide additional prediction.

A simple example will illustrate these relations. Suppose that two observers rate two stimuli with perfect consistency, but one observer uses the numbers "1" and "2", while the other uses the numbers "3" and "4", for the first and second stimuli, respectively. Now if we know the response we can perfectly predict the stimulus, and it is no help to us to know which observer uses which responses. If we use the stimulus to predict the response, however, we must first know which observer is involved to know which response is used. On the other hand, suppose that both observers use the numbers "1" and "2", but in reverse relation to the two stimuli. Knowing the stimulus alone is now of no value in predicting the reponse, or vice versa.

However, if knowledge of the observer is added, then perfect prediction of the response from the stimulus or of the stimulus from the response is possible, due to the fact that the addition of observers as a predictor variable adds an interaction term in which all of the available prediction lies.

As a concrete example, these two cases are extreme and serve only to illustrate the principles. One fact remains, however, which is that, in such a case, prediction of stimuli can never be better than prediction of responses, and will normally be somewhat poorer.

This type of problem can be carried even further. In the same experiment, Garner actually determined the multiple contingent uncertainty when three predictor variables were used, the fourth variable being the stimulus which preceded the one being judged ($P$). Furthermore, in the experimental design, all pairs of stimuli were made to occur equally often, so that $P$ and $S$ were orthogonally related. Now the two types of prediction are

$$U(R:S,O,P) = U(R:S:O:P) \\ - U(S:O:P) \quad [16a]$$

and

$$U(S:R,O,P) = U(R:S:O:P) \\ - U(R:O:P) \quad [16b]$$

following the form of Equation 12. In this case, when $R$ is predicted from $S$, $O$, and $P$, the multiple contingent uncertainty is the total contingent uncertainty in the four-variable matrix minus the total contingent uncertainty in the three-variable matrix involving the three predictor variables. Since these three predictor variables were all orthogonal in the experimental design, this term is zero, and this multiple contingent uncertainty contains all of the contingent uncertainty available in the matrix.

When, however, $S$ is predicted from the other three variables, the orthogonal condition does not hold, since of the three variables only $O$ and $P$ are orthogonal. This term will thus be positive, and will be fairly substantial, since it includes contingent uncertainties between $R$ and $O$, and between $R$ and $P$, as well as the three-variable interaction which must (in this case) also be positive. Thus, there would be substantial error in assuming that these two multiple contingent uncertainties are the same.

## Redundancy of Printed English

One common application of information theory and uncertainty analysis is to situations where events occur in a sequence which is not completely random, i.e., in which there are sequential dependencies. Much of human behavior is concerned with making predictions about events which are unknown, or only partially known, on the basis of sequentially provided information. Language is a very good example of such an event series in which there exist sequential dependencies, and several studies have been done on printed English to determine how much predictability is available from sequences of letters. The term "redundancy" has been used to describe the amount of uncertainty in a given letter which is predictable on the basis of other known letters. The equations described above point out some interesting aspects of redundancy of printed English.

When we do an analysis of sequences of letters, we form a data matrix in which the series of letters itself is one variable, and the series displaced by one or more units constitute the other variables. For present

purposes we shall refer to the first order series as variable $w$, the series displaced one step as variable $x$, and the series displaced two steps as variable $y$, and we will be dealing with a three-dimensional matrix. In this special three-variable case there are some restrictions on the contingent uncertainties which exist. For example, we know that the term $U(w:x)$ is identical to the term $U(x:y)$, since in both cases we are describing the contingent uncertainty for a one-step displacement of the series. In addition, the contingent uncertainty $U(w:y)$ is less than either of the others, since that is the contingency for a two-step displacement, and a letter cannot be predicted as well from one two steps behind as from the immediately preceding one.

Now we know from Equation 11 that the multiple contingent uncertainty for predicting $x$, the middle letter of a series of three, must be greater than that for predicting either $w$ or $y$ from the other two, since when $x$ is being predicted, the contingent uncertainty between the two predictor variables, $w$ and $y$, is less than the contingent uncertainties between the other pairs of predictor variables. In other words, prediction of the middle letter of a sequence of three must be better than prediction of either end of the sequence on the basis of the other two. Similar relations hold for longer sequences of letters, and we can make the general statement that if any fixed number of letters is available for prediction, the best prediction will occur if the letter being predicted is in the middle of the total sequence rather than at either end. Miller and Friedman (6) have experimentally shown this to be true for sequences of 5, 7, and 11 letters.

Newman and Gerstman (7) have provided data which allow us to see just how much better prediction of middle letters should be, compared to prediction of end letters. Their data show, for the three-letter case, that the contingent uncertainty for a one-step displacement is 0.91 bit, and for a two-step displacement it is 0.42 bit. This difference is 0.49 bit, and reference to Equation 11 shows that this is the exact amount by which prediction of the middle of three letters will be better than prediction of either end. This is a substantial difference.

As pointed out above, the situation becomes more complicated when we deal with longer letter sequences because of the greater number of interaction terms. However, we can assume that these terms are negligible, and determine the multiple contingent uncertainty for predicting omitted letters from other length sequences simply by adding up the various two-term contingent uncertainties involved in the multiple contingent uncertainty. Using again the Newman and Gerstman data, for an eleven-letter sequence, the contingent uncertainty for predicting either end letter on the basis of the other ten is 2.17 bits, while that for predicting the middle letter, on the basis of five letters on each side, is 3.72 bits. For seven-letter sequences, prediction of the end letters gives a multiple contingent uncertainty of 1.95 bits, while prediction of the middle letter gives 3.18 bits. These figures compare quite well with those found by Miller and Friedman, although the agreement is better for middle letters than for ends. It seems, from their data, that humans make more effective use of information presented symmetrically than they do of unilateral information, since their data show higher residual uncertainties for the ends than these calculations predict.

These relations point out the need

for a more accurate interpretation of the meaning of the term redundancy. This term has usually been defined for prediction of letters on the basis of preceding letters only, and the figure of 50 per cent redundancy is accepted as a reasonable approximation. Such a figure, however, can be interpreted quantitatively and accurately only when it is applied to this particular prediction situation. We cannot infer from it how many letters in printed English can be omitted and still allow perfect reconstruction. Clearly, with a figure of 50 per cent redundancy for one direction only, we cannot expect perfect reproducibility unless we have this amount available on both sides of the item being predicted. In other words, we need 100 per cent redundancy for perfect prediction, and this can be obtained only when a single letter is omitted from a very long sequence, so that maximum redundancy is available on both sides of the letter being predicted. Perhaps our thinking on this problem has been limited by our natural tendency to deal with antecedent and consequent events; but we can think of both antecedent and subsequent events as determining the consequent events, and combinations of these certainly are more effective than either alone.

### Response to a Continuing Series

A type of behavior for which the relations shown here have important implications is that in which a continuing series of events occurs and a human must make a differential response to each event as it occurs. Such tasks as copying telegraphic code, taking dictation, reporting target positions on a radar, and target tracking are all of this type.

To talk about a simplified task of this sort, let us assume that an opera-

tor is read a series of numbers and he is required to push a different button for each different number he hears, much as he would if he were adding up a column of figures on a machine as they were being read to him. The operator, of course, is trying to carry out his task with minimum error, so that he wants to get all the information possible about what number was read to him. Initially, let us suppose also that the numbers do not occur in a completely random order, but that each number is to some extent predictable from the preceding ones in a fairly simple way. Such a series, of course, will show sequential dependencies, and, with rare exceptions, the contingent uncertainties between the series and its displacements will be largest for adjacent numbers and will decrease as the number series is displaced from itself more and more. Now we know from the relations shown in this paper that the maximum predictability for any single number will be obtained if the operator is able to use symmetrical prediction of it, i.e., to use both forward and backward prediction. In order to use both forward and backward prediction, however, he must delay his response to a particular number until he has heard some subsequent numbers. He must, in other words, lag behind the presented series. This is, of course, a phenomenon commonly observed in everyday life.

Suppose we want to determine the optimum lag which such an operator should use. We can make some reasonable predictions, but these predictions will depend on a number of factors.

1. *The memory function.* It is clear that the operator cannot delay his response so long that he cannot remember all of the items which are potentially useful for prediction of any one

TABLE 1

ILLUSTRATION OF EFFECT OF MEMORY AND CONTINGENT UNCERTAINTIES ON OPTIMUM LAG

| Unit Step | Contingent Uncertainty | Lag of 2 | | Lag of 3 | | Lag of 1 | |
|---|---|---|---|---|---|---|---|
| | | Weight | Weighted Value | Weight | Weighted Value | Weight | Weighted Value |
| −3 | .30 | | | | | .2 | .06 |
| −2 | .50 | .2 | .10 | | | .4 | .20 |
| −1 | 1.00 | .4 | .40 | .2 | .20 | .6 | .60 |
| 0 | | .6 | | .4 | | .8 | |
| +1 | 1.00 | .8 | .80 | .6 | .60 | 1.0 | 1.00 |
| +2 | .50 | 1.0 | .50 | .8 | .40 | | |
| +3 | .30 | | | 1.0 | .30 | | |
| Σ weighted values | | | 1.80 | | 1.50 | | 1.86 |

item. Let us assume that a reasonable memory span for such a task is five units. Now if all five units can be remembered equally well, the optimum lag for the subject would be two units behind the one to which the response is being made, since this lag will give two units on either side of the one in question, with the optimum of completely symmetrical prediction.

Actually, it is unrealistic to assume that all of the items can be remembered equally well. It seems more reasonable to assume that the items last heard will be more clearly remembered, and thus that the contingent uncertainty available from them will be more useful. We can illustrate the effects of such a memory function by using a weighting function which gives greater value to recent items than to remote items. Table 1 contains a set of purely hypothetical numbers to illustrate how such a memory factor would affect the optimum lag. The first column shows the unit step for each item, with "O" being the item to which the response is to be made, negative numbers referring to items which occurred before that one, and positive numbers being the later, more recent items. Column 2 shows the contingent uncertainty for each displacement with respect to the "O" item. Column 3 shows the weights which have been assigned to each item being used for predictive purposes, with the largest weight for the most recent item. Column 4 shows the resultant weighted value of the contingent uncertainty for each step, with the total at the bottom. For these two columns, we have assumed a lag of two steps. The next two columns show the same calculations for a lag of three steps, and the last two columns for a lag of one step.

With an equiweighted function, the maximum multiple contingent uncertainty occurs with a lag of two, and has a value of 3.00 bits (i.e., all weights are 1.0). When the differential weights are used, however, the maximum-weighted contingent uncertainty occurs with a lag of one, and has a value of 1.86 bits. Thus, if we assume a memory differential for the various steps, two things happen: the total contingent uncertainty decreases, and the optimum lag also decreases. In general, where responses are made to a continued time series, we would expect the optimum lag to be some-

what less than that which gives complete symmetry to the predictors.

2. *The amount of redundancy.* A second factor which affects the optimum lag is the total amount of redundancy in the series, as well as its distribution over the various steps. This point need not be labored. A lag which gains no predictability can be of no value. Clearly if the series is completely random (no redundancy), there is no advantage to the operator in lagging because he gains no redundancy or information when he does. Similarly, if only one item on either side gives any predictability, then no lag beyond one step is ever of any value.

3. *The amount of noise.* One last factor which will affect the optimum lag for the operator is the amount of noise whose effects have to be offset. In our illustration, the term noise can refer to acoustical noise, or in the more general case, it can refer to any factor which decreases the probability of a correct perception of the item. The importance of noise is that it determines the extent to which the operator needs the predictability which he can obtain from the series redundancy. Suppose, for example, that the task is being carried out under ideal circumstances, where each number can be heard quite clearly. Under these circumstances, the operator can respond immediately with essentially perfect accuracy, and it would make little difference to him whether the series is partially redundant or not. In either case he would not lag, since there is nothing to be gained by it. If, on the other hand, there is sufficient noise to make the perception imperfect, then the operator will frequently find himself unsure of the proper response, and will need some additional redundancy to improve the probability of his making a correct response.

Thus, in summary, we can assume that with this type of task operators will lag if there is some noise in the situation and if the series has sequential dependencies in it. If these conditions are satisfied, then the exact amount of the lag will depend on his memory for the items presented and on the distribution of the redundancy in the series of events.

## Summary

In this paper we have presented some equations originating in information theory which show some valuable properties of uncertainty analysis. These equations are presented in a form which emphasizes the fact that an uncertainty analysis can be carried out so that all variables involved can be dealt with in a completely symmetric manner. It is possible, for example, to talk about a single number which represents the total contingent uncertainty in a multidimensional matrix, and then to demonstrate how parts or all of this contingent uncertainty can be made available for prediction of just one of the variables. From such equations it is possible to demonstrate certain relations which must hold between prediction of different variables in a single matrix.

These relations have important implications for several areas of psychology, either in clarifying the meaning of terms used in statistical analysis or in showing how behavior ought to occur if it is to maximize the use of information. As one illustration, it is shown that information transmission in the multidimensional case cannot be as good when stimuli are predicted as when responses are predicted. Another illustration shows that the usual measure of redundancy of printed English can be interpreted accurately only when prediction is

made unilaterally, but that maximum predictability of single letters occurs when prediction (on the basis of redundancy) is allowed to operate in both the forward and backward directions. A third illustration shows that when human operators have to respond to a continuing series which is nonrandom, optimum performance (with an error criterion) will occur only when the operator lags behind the series to which he is responding; furthermore, the optimum lag will be a function of the memory of the operator, the amount of noise in the situation, and the distribution of redundancy throughout the series.

These examples are in no sense inclusive, and the implications of equations presented here range into many areas of psychology. These illustrations are sufficient, however, to point out several possible directions for research, and to clarify some concepts now in use in psychology.

## REFERENCES

1. GARNER, W. R. An informational analysis of absolute judgments of loudness. *J. exp. Psychol.*, 1953, **46**, 373–380.
2. GARNER, W. R., & HAKE, H. W. The amount of information in absolute judgments. *Psychol. Rev.*, 1951, **58**, 446–459.
3. GARNER, W. R., & McGILL, W. J. The relation between information and variance analyses. *Psychometrika*, 1956, **21**, 219–228.
4. McGILL, W. J. Multivariate information transmission. *Psychometrika*, 1954, **19**, 97–116.
5. McGILL, W. J., & QUASTLER, H. Standardized nomenclature. In H. Quastler (Ed.), *Information theory in psychology*. Glencoe, Ill.: Free Press, 1955. Pp. 83–92.
6. MILLER, G. A., & FRIEDMAN, E. A. The reconstruction of mutilated English texts. *Information and Control*, 1957, **1**, 38–55.
7. NEWMAN, E. B., & GERSTMAN, L. J. A new method for analyzing printed English. *J. exp. Psychol.*, 1952, **44**, 114–125.
8. SHANNON, C. E. A mathematical theory of communication. *Bell Syst. tech. J.*, 1948, **27**, 379–423, 623–656.

# HEREDITY, ENVIRONMENT, AND THE QUESTION "HOW?"[1]

ANNE ANASTASI

*Fordham University*

Two or three decades ago, the so-called heredity-environment question was the center of lively controversy. Today, on the other hand, many psychologists look upon it as a dead issue. It is now generally conceded that both hereditary and environmental factors enter into all behavior. The reacting organism is a product of its genes and its past environment, while present environment provides the immediate stimulus for current behavior. To be sure, it can be argued that, although a given trait may result from the combined influence of hereditary and environmental factors, a specific difference in this trait between individuals or between groups may be traceable to either hereditary or environmental factors alone. The design of most traditional investigations undertaken to identify such factors, however, has been such as to yield inconclusive answers. The same set of data has frequently led to opposite conclusions in the hands of psychologists with different orientations.

Nor have efforts to determine the proportional contribution of hereditary and environmental factors to observed individual differences in given traits met with any greater success. Apart from difficulties in controlling conditions, such investigations have usually been based upon the implicit assumption that hereditary and environmental factors combine in an additive fashion. Both geneticists and psychologists have repeatedly demonstrated, however, that a more tenable hypothesis is that of interaction (15, 22, 28, 40). In other words, the

nature and extent of the influence of each type of factor depend upon the contribution of the other. Thus the proportional contribution of heredity to the variance of a given trait, rather than being a constant, will vary under different environmental conditions. Similarly, under different hereditary conditions, the relative contribution of environment will differ. Studies designed to estimate the proportional contribution of heredity and environment, however, have rarely included measures of such interaction. The only possible conclusion from such research would thus seem to be that both heredity and environment contribute to all behavior traits and that the extent of their respective contributions cannot be specified for any trait. Small wonder that some psychologists regard the heredity-environment question as unworthy of further consideration!

But is this really all we can find out about the operation of heredity and environment in the etiology of behavior? Perhaps we have simply been asking the wrong questions. The traditional questions about heredity and environment may be intrinsically unanswerable. Psychologists began by asking *which* type of factor, hereditary or environmental, is responsible for individual differences in a given trait. Later, they tried to discover *how much* of the variance was attributable to heredity and how much to environment. It is the primary contention of this paper that a more fruitful approach is to be found in the question *"How?"* There is still much to be learned about the specific *modus operandi* of hereditary and environmental factors in the development of behavioral

differences. And there are several current lines of research which offer promising techniques for answering the question "How?"

VARIETY OF INTERACTION MECHANISMS

*Hereditary factors.* If we examine some of the specific ways in which hereditary factors may influence behavior, we cannot fail but be impressed by their wide diversity. At one extreme, we find such conditions as phenylpyruvic amentia and amaurotic idiocy. In these cases, certain essential physical prerequisites for normal intellectual development are lacking as a result of hereditary metabolic disorders. In our present state of knowledge, there is no environmental factor which can completely counteract this hereditary deficit. The individual will be mentally defective, regardless of the type of environmental conditions under which he is reared.

A somewhat different situation is illustrated by hereditary deafness, which may lead to intellectual retardation through interference with normal social interaction, language development, and schooling. In such a case, however, the hereditary handicap can be offset by appropriate adaptations of training procedures. It has been said, in fact, that the degree of intellectual backwardness of the deaf is an index of the state of development of special instructional facilities. As the latter improve, the intellectual retardation associated with deafness is correspondingly reduced.

A third example is provided by inherited susceptibility to certain physical diseases, with consequent protracted ill health. If environmental conditions are such that illness does in fact develop, a number of different behavioral effects may follow. Intellectually, the individual may be handicapped by his inability to attend school regularly. On the other hand, depending upon age of onset, home conditions, parental status, and similar factors, poor health may have the effect of concentrating the individual's energies upon intellectual pursuits. The curtailment of participation in athletics and social functions may serve to strengthen interest in reading and other sedentary activities. Concomitant circumstances would also determine the influence of such illness upon personality development. And it is well known that the latter effects could run the gamut from a deepening of human sympathy to psychiatric breakdown.

Finally, heredity may influence behavior through the mechanism of social stereotypes. A wide variety of inherited physical characteristics have served as the visible cues for identifying such stereotypes. These cues thus lead to behavioral restrictions or opportunities and—at a more subtle level —to social attitudes and expectancies. The individual's own self concept tends gradually to reflect such expectancies. All of these influences eventually leave their mark upon his abilities and inabilities, his emotional reactions, goals, ambitions, and outlook on life.

The geneticist Dobzhansky illustrates this type of mechanism by means of a dramatic hypothetical situation. He points out that, if there were a culture in which the carriers of blood group AB were considered aristocrats and those of blood group O laborers, then the blood-group genes would become important hereditary determiners of behavior (12, p. 147). Obviously the association between blood group and behavior would be specific to that culture. But such specificity is an essential property of the causal mechanism under consideration.

More realistic examples are not hard to find. The most familiar instances occur in connection with constitutional types, sex, and race. Sex and skin pig-

mentation obviously depend upon heredity. General body build is strongly influenced by hereditary components, although also susceptible to environmental modification. That all these physical characteristics may exert a pronounced effect upon behavior within a given culture is well known. It is equally apparent, of course, that in different cultures the behavioral correlates of such hereditary physical traits may be quite unlike. A specific physical cue may be completely unrelated to individual differences in psychological traits in one culture, while closely correlated with them in another. Or it may be associated with totally dissimilar behavior characteristics in two different cultures.

It might be objected that some of the illustrations which have been cited do not properly exemplify the operation of hereditary mechanisms in behavior development, since hereditary factors enter only indirectly into the behavior in question. Closer examination, however, shows this distinction to be untenable. First it may be noted that the influence of heredity upon behavior is always indirect. No psychological trait is ever inherited as such. All we can ever say directly from behavioral observations is that a given trait shows evidence of being influenced by certain "inheritable unknowns." This merely defines a problem for genetic research; it does not provide a causal explanation. Unlike the blood groups, which are close to the level of primary gene products, psychological traits are related to genes by highly indirect and devious routes. Even the mental deficiency associated with phenylketonuria is several steps removed from the chemically defective genes that represent its hereditary basis. Moreover, hereditary influences cannot be dichotomized into the more direct and the less direct. Rather do they represent a whole "continuum of indirectness," along which are found all degrees of remoteness of causal links. The examples already cited illustrate a few of the points on this continuum.

It should be noted that as we proceed along the continuum of indirectness, the range of variation of possible outcomes of hereditary factors expands rapidly. At each step in the causal chain, there is fresh opportunity for interaction with other hereditary factors as well as with environmental factors. And since each interaction in turn determines the direction of subsequent interactions, there is an ever-widening network of possible outcomes. If we visualize a simple sequential grid with only two alternatives at each point, it is obvious that there are two possible outcomes in the one-stage situation, four outcomes at the second stage, eight at the third, and so on in geometric progression. The actual situation is undoubtedly much more complex, since there will usually be more than two alternatives at any one point.

In the case of the blood groups, the relation to specific genes is so close that no other concomitant hereditary or environmental conditions can alter the outcome. If the organism survives at all, it will have the blood group determined by its genes. Among psychological traits, on the other hand, some variation in outcome is always possible as a result of concurrent circumstances. Even in cases of phenylketonuria, intellectual development will exhibit some relationship with the type of care and training available to the individual. That behavioral outcomes show progressive diversification as we proceed along the continuum of indirectness is brought out by the other examples which were cited. Chronic illness *can* lead to scholarly renown or to intellectual immaturity; a mesomorphic physique *can* be a contributing factor in juvenile delinquency or in the at-

tainment of a college presidency! Published data on Sheldon somatotypes provide some support for both of the latter outcomes.

Parenthetically, it may be noted that geneticists have sometimes used the term "norm of reaction" to designate the range of variation of possible outcomes of gene properties (cf. 13, p. 161). Thus heredity sets the "norm" or limits within which environmental differences determine the eventual outcome. In the case of some traits, such as blood groups or eye color, this norm is much narrower than in the case of other traits. Owing to the rather different psychological connotations of both the words "norm" and "reaction," however, it seems less confusing to speak of the "range of variation" in this context.

A large portion of the continuum of hereditary influences which we have described coincides with the domain of somatopsychological relations, as defined by Barker et al. (6). Under this heading, Barker includes "variations in physique that affect the psychological situation of a person by influencing the effectiveness of his body as a tool for actions or by serving as a stimulus to himself or others" (6, p. 1). Relatively direct neurological influences on behavior, which have been the traditional concern of physiological psychology, are excluded from this definition, Barker being primarily concerned with what he calls the "social psychology of physique." Of the examples cited in the present paper, deafness, severe illness, and the physical characteristics associated with social stereotypes would meet the specifications of somatopsychological factors.

The somatic factors to which Barker refers, however, are not limited to those of hereditary origin. Bodily conditions attributable to environmental causes operate in the same sorts of somatopsychological relations as those traceable to heredity. In fact, heredity-environment distinctions play a minor part in Barker's approach.

*Environmental factors: organic.* Turning now to an analysis of the role of environmental factors in behavior, we find the same etiological mechanisms which were observed in the case of hereditary factors. First, however, we must differentiate between two classes of environmental influences: (a) those producing organic effects which may in turn influence behavior and (b) those serving as direct stimuli for psychological reactions. The former may be illustrated by food intake or by exposure to bacterial infection; the latter, by tribal initiation ceremonies or by a course in algebra. There are no completely satisfactory names by which to designate these two classes of influences. In an earlier paper by Anastasi and Foley (4), the terms "structural" and "functional" were employed. However, "organic" and "behavioral" have the advantage of greater familiarity in this context and may be less open to misinterpretation. Accordingly, these terms will be used in the present paper.

Like hereditary factors, environmental influences of an organic nature can also be ordered along a continuum of indirectness with regard to their relation to behavior. This continuum closely parallels that of hereditary factors. One end is typified by such conditions as mental deficiency resulting from cerebral birth injury or from prenatal nutritional inadequacies. A more indirect etiological mechanism is illustrated by severe motor disorder—as in certain cases of cerebral palsy—*without* accompanying injury to higher neurological centers. In such instances, intellectual retardation may occur as an indirect result of the motor handicap, through the curtailment of educational and social activities. Obviously this causal mechanism

corresponds closely to that of hereditary deafness cited earlier in the paper.

Finally, we may consider an environmental parallel to the previously discussed social stereotypes which were mediated by hereditary physical cues. Let us suppose that a young woman with mousy brown hair becomes transformed into a dazzling golden blonde through environmental techniques currently available in our culture. It is highly probable that this metamorphosis will alter, not only the reactions of her associates toward her, but also her own self concept and subsequent behavior. The effects could range all the way from a rise in social poise to a drop in clerical accuracy!

Among the examples of environmentally determined organic influences which have been described, all but the first two fit Barker's definition of somatopsychological factors. With the exception of birth injuries and nutritional deficiencies, all fall within the social psychology of physique. Nevertheless, the individual factors exhibit wide diversity in their specific *modus operandi* —a diversity which has important practical as well as theoretical implications.

*Environmental factors: behavioral.* The second major class of environmental factors—the behavioral as contrasted to the organic—are by definition direct influences. The immediate effect of such environmental factors is always a behavioral change. To be sure, some of the initial behavioral effects may themselves indirectly affect the individual's later behavior. But this relationship can perhaps be best conceptualized in terms of breadth and permanence of effects. Thus it could be said that we are now dealing, not with a continuum of indirectness, as in the case of hereditary and organic-environmental factors, but rather with a continuum of breadth.

Social class membership may serve as an illustration of a relatively broad, pervasive, and enduring environmental factor. Its influence upon behavior development may operate through many channels. Thus social level may determine the range and nature of intellectual stimulation provided by home and community through books, music, art, play activities, and the like. Even more far-reaching may be the effects upon interests and motivation, as illustrated by the desire to perform abstract intellectual tasks, to surpass others in competitive situations, to succeed in school, or to gain social approval. Emotional and social traits may likewise be influenced by the nature of interpersonal relations characterizing homes at different socioeconomic levels. Somewhat more restricted in scope than social class, although still exerting a relatively broad influence, is amount of formal schooling which the individual is able to obtain.

A factor which may be wide or narrow in its effects, depending upon concomitant circumstances, is language handicap. Thus the bilingualism of an adult who moves to a foreign country with inadequate mastery of the new language represents a relatively limited handicap which can be readily overcome in most cases. At most, the difficulty is one of communication. On the other hand, some kinds of bilingualism in childhood may exert a retarding influence upon intellectual development and may under certain conditions affect personality development adversely (2, 5, 10). A common pattern in the homes of immigrants is that the child speaks one language at home and another in school, so that his knowledge of each language is limited to certain types of situations. Inadequate facility with the language of the school interferes with the acquisition of basic concepts, intellectual skills, and information. The frustration engendered by scholastic difficulties may in turn lead to discouragement and general dis-

like of school. Such reactions can be found, for example, among a number of Puerto Rican children in New York City schools (3). In the case of certain groups, moreover, the child's foreign language background may be perceived by himself and his associates as a symbol of minority group status and may thereby augment any emotional maladjustment arising from such status (34).

A highly restricted environmental influence is to be found in the opportunity to acquire specific items of information occurring in a particular intelligence test. The fact that such opportunities may vary with culture, social class, or individual experiential background is at the basis of the test user's concern with the problem of coaching and with "culture-free" or "culture-fair" tests (cf. 1, 2). If the advantage or disadvantage which such experiential differences confer upon certain individuals is strictly confined to performance on the given test, it will obviously reduce the validity of the test and should be eliminated.

In this connection, however, it is essential to know the breadth of the environmental influence in question. A fallacy inherent in many attempts to develop culture-fair tests is that the breadth of cultural differentials is not taken into account. Failure to consider breadth of effect likewise characterizes certain discussions of coaching. If, in coaching a student for a college admission test, we can improve his knowledge of verbal concepts and his reading comprehension, he will be better equipped to succeed in college courses. His performance level will thus be raised, not only on the test, but also on the criterion which the test is intended to predict. To try to devise a test which is not susceptible to such coaching would merely reduce the effectiveness of the test. Similarly, efforts to rule out cultural differentials from test items so as to make them equally "fair" to subjects in different social classes or in different cultures may merely limit the usefulness of the test, since the same cultural differentials may operate within the broader area of behavior which the test is designed to sample.

## METHODOLOGICAL APPROACHES

The examples considered so far should suffice to highlight the wide variety of ways in which hereditary and environmental factors may interact in the course of behavior development. There is clearly a need for identifying explicitly the etiological mechanism whereby any given hereditary or environmental condition ultimately leads to a behavioral characteristic—in other words, the "how" of heredity and environment. Accordingly, we may now take a quick look at some promising methodological approaches to the question "how."

Within the past decade, an increasing number of studies have been designed to trace the connection between specific factors in the hereditary backgrounds or in the reactional biographies of individuals and their observed behavioral characteristics. There has been a definite shift away from the predominantly descriptive and correlational approach of the earlier decades toward more deliberate attempts to verify explanatory hypotheses. Similarly, the cataloguing of group differences in psychological traits has been giving way gradually to research on *changes* in group characteristics following altered conditions.

Among recent methodological developments, we have chosen seven as being particularly relevant to the analysis of etiological mechanisms. The first represents an extension of selective breeding investigations to permit the identification of specific hereditary conditions underlying the observed behavioral differences. When early selective breeding investigations such as those of Tryon

(36) on rats indicated that "maze learning ability" was inherited, we were still a long way from knowing what was actually being transmitted by the genes. It was obviously not "maze learning ability" as such. Twenty—or even ten —years ago, some psychologists would have suggested that it was probably general intelligence. And a few might even have drawn a parallel with the inheritance of human intelligence.

But today investigators have been asking: Just what makes one group of rats learn mazes more quickly than the other? Is it differences in motivation, emotionality, speed of running, general activity level? If so, are these behavioral characteristics in turn dependent upon group differences in glandular development, body weight, brain size, biochemical factors, or some other organic conditions? A number of recent and ongoing investigations indicate that attempts are being made to trace, at least part of the way, the steps whereby certain chemical properties of the genes may ultimately lead to specified behavior characteristics.

An example of such a study is provided by Searle's (31) follow-up of Tryon's research. Working with the strains of maze-bright and maze-dull rats developed by Tryon, Searle demonstrated that the two strains differed in a number of emotional and motivational factors, rather than in ability. Thus the strain differences were traced one step further, although many links still remain to be found between maze learning and genes. A promising methodological development within the same general area is to be found in the recent research of Hirsch and Tryon (18). Utilizing a specially devised technique for measuring individual differences in behavior among lower organisms, these investigators launched a series of studies on selective breeding for behavioral characteristics in the fruit fly, Dro-sophila. Such research can capitalize on the mass of available genetic knowledge regarding the morphology of Dro-sophila, as well as on other advantages of using such an organism in genetic studies.

Further evidence of current interest in the specific hereditary factors which influence behavior is to be found in an extensive research program in progress at the Jackson Memorial Laboratory, under the direction of Scott and Fuller (30). In general, the project is concerned with the behavioral characteristics of various breeds and cross-breeds of dogs. Analyses of some of the data gathered to date again suggest that "differences in performance are produced by differences in emotional, motivational, and peripheral processes, and that genetically caused differences in central processes may be either slight or non-existent" (29, p. 225). In other parts of the same project, breed differences in physiological characteristics, which may in turn be related to behavioral differences, have been established.

A second line of attack is the exploration of possible relationships between behavioral characteristics and physiological variables which may in turn be traceable to hereditary factors. Research on EEG, autonomic balance, metabolic processes, and biochemical factors illustrates this approach. A lucid demonstration of the process of tracing a psychological condition to genetic factors is provided by the identification and subsequent investigation of phenylpyruvic amentia. In this case, the causal chain from defective gene, through metabolic disorder and consequent cerebral malfunctioning, to feeblemindedness and other overt symptoms can be described step by step (cf. 32; 33, pp. 389–391). Also relevant are the recent researches on neurological and biochemical correlates of schizo-

phrenia (9). Owing to inadequate methodological controls, however, most of the findings of the latter studies must be regarded as tentative (19).

Prenatal environmental factors provide a third avenue of fruitful investigation. Especially noteworthy is the recent work of Pasamanick and his associates (27), which demonstrated a tie-up between socioeconomic level, complications of pregnancy and parturition, and psychological disorders of the offspring. In a series of studies on large samples of whites and Negroes in Baltimore, these investigators showed that various prenatal and paranatal disorders are significantly related to the occurrence of mental defect and psychiatric disorders in the child. An important source of such irregularities in the process of childbearing and birth is to be found in deficiencies of maternal diet and in other conditions associated with low socioeconomic status. An analysis of the data did in fact reveal a much higher frequency of all such medical complications in lower than in higher socioeconomic levels, and a higher frequency among Negroes than among whites.

Direct evidence of the influence of prenatal nutritional factors upon subsequent intellectual development is to be found in a recent, well controlled experiment by Harrell et al. (16). The subjects were pregnant women in low-income groups, whose normal diets were generally quite deficient. A dietary supplement was administered to some of these women during pregnancy and lactation, while an equated control group received placebos. When tested at the ages of three and four years, the offspring of the experimental group obtained a significantly higher mean IQ than did the offspring of the controls.

Mention should also be made of animal experiments on the effects of such factors as prenatal radiation and neo-natal asphyxia upon cerebral anomalies as well as upon subsequent behavior development. These experimental studies merge imperceptibly into the fourth approach to be considered, namely, the investigation of the influence of early experience upon the eventual behavioral characteristics of animals. Research in this area has been accumulating at a rapid rate. In 1954, Beach and Jaynes (8) surveyed this literature for the *Psychological Bulletin,* listing over 130 references. Several new studies have appeared since that date (e.g., 14, 21, 24, 25, 35). The variety of factors covered ranges from the type and quantity of available food to the extent of contact with human culture. A large number of experiments have been concerned with various forms of sensory deprivation and with diminished opportunities for motor exercise. Effects have been observed in many kinds of animals and in almost all aspects of behavior, including perceptual responses, motor activity, learning, emotionality, and social reactions.

In their review, Beach and Jaynes pointed out that research in this area has been stimulated by at least four distinct theoretical interests. Some studies were motivated by the traditional concern with the relative contribution of maturation and learning to behavior development. Others were designed in an effort to test certain psychoanalytic theories regarding infantile experiences, as illustrated by studies which limited the feeding responses of young animals. A third relevant influence is to be found in the work of the European biologist Lorenz (23) on early social stimulation of birds, and in particular on the special type of learning for which the term "imprinting" has been coined. A relatively large number of recent studies have centered around Hebb's (17) theory regarding the importance of early perceptual experiences upon subsequent

performance in learning situations. All this research represents a rapidly growing and promising attack on the *modus operandi* of specific environmental factors.

The human counterpart of these animal studies may be found in the comparative investigation of child-rearing practices in different cultures and subcultures. This represents the fifth approach in our list. An outstanding example of such a study is that by Whiting and Child (38), published in 1953. Utilizing data on 75 primitive societies from the Cross-Cultural Files of the Yale Institute of Human Relations, these investigators set out to test a number of hypotheses regarding the relationships between child-rearing practices and personality development. This analysis was followed up by field observations in five cultures, the results of which have not yet been reported (cf. 37).

Within our own culture, similar surveys have been concerned with the diverse psychological environments provided by different social classes (11). Of particular interest are the study by Williams and Scott (39) on the association between socioeconomic level, permissiveness, and motor development among Negro children, and the exploratory research by Milner (26) on the relationship between reading readiness in first-grade children and patterns of parent-child interaction. Milner found that upon school entrance the lower-class child seems to lack chiefly two advantages enjoyed by the middle-class child. The first is described as "a warm positive family atmosphere or adult-relationship pattern which is more and more being recognized as a motivational prerequisite of any kind of adult-controlled learning." The lower-class children in Milner's study perceived adults as predominantly hostile. The second advantage is an extensive opportunity to interact verbally with adults in the family. The latter point is illustrated by parental attitudes toward mealtime conversation, lower-class parents tending to inhibit and discourage such conversation, while middle-class parents encourage it.

Most traditional studies on child-rearing practices have been designed in terms of a psychoanalytic orientation. There is need for more data pertaining to other types of hypotheses. Findings such as those of Milner on opportunities for verbalization and the resulting effects upon reading readiness represent a step in this direction. Another possible source of future data is the application of the intensive observational techniques of psychological ecology developed by Barker and Wright (7) to widely diverse socioeconomic groups.

A sixth major approach involves research on the previously cited somatopsychological relationships (6). To date, little direct information is available on the precise operation of this class of factors in psychological development. The multiplicity of ways in which physical traits—whether hereditary or environmental in origin—may influence behavior thus offers a relatively unexplored field for future study.

The seventh and final approach to be considered represents an adaptation of traditional twin studies. From the standpoint of the question "How?" there is need for closer coordination between the usual data on twin resemblance and observations of the family interactions of twins. Available data already suggest, for example, that closeness of contact and extent of environmental similarity are greater in the case of monozygotic than in the case of dizygotic twins (cf. 2). Information on the social reactions of twins toward each other and the specialization of roles is likewise of interest (2). Especially useful would be longitudinal stud-

ies of twins, beginning in early infancy and following the subjects through school age. The operation of differential environmental pressures, the development of specialized roles, and other environmental influences could thus be more clearly identified and correlated with intellectual and personality changes in the growing twins.

Parenthetically, I should like to add a remark about the traditional applications of the twin method, in which persons in different degrees of hereditary and environmental relationships to each other are simply compared for behavioral similarity. In these studies, attention has been focused principally upon the amount of resemblance of monozygotic as contrasted to dizygotic twins. Yet such a comparison is particularly difficult to interpret because of the many subtle differences in the environmental situations of the two types of twins. A more fruitful comparison would seem to be that between dizygotic twins and siblings, for whom the hereditary similarity is known to be the same. In Kallmann's monumental research on psychiatric disorders among twins (20), for example, one of the most convincing bits of evidence for the operation of hereditary factors in schizophrenia is the fact that the degrees of concordance for dizygotic twins and for siblings were practically identical. In contrast, it will be recalled that in intelligence test scores dizygotic twins resemble each other much more closely than do siblings—a finding which reveals the influence of environmental factors in intellectual development.

## SUMMARY

The heredity-environment problem is still very much alive. Its viability is assured by the gradual replacement of the questions, "Which one?" and "How much?" by the more basic and appropriate question, "How?" Hereditary in-

fluences—as well as environmental factors of an organic nature—vary along a "continuum of indirectness." The more indirect their connection with behavior, the wider will be the range of variation of possible outcomes. One extreme of the continuum of indirectness may be illustrated by brain damage leading to mental deficiency; the other extreme, by physical characteristics associated with social stereotypes. Examples of factors falling at intermediate points include deafness, physical diseases, and motor disorders. Those environmental factors which act directly upon behavior can be ordered along a continuum of breadth or permanence of effect, as exemplified by social class membership, amount of formal schooling, language handicap, and familiarity with specific test items.

Several current lines of research offer promising techniques for exploring the *modus operandi* of hereditary and environmental factors. Outstanding among them are investigations of: (a) hereditary conditions which underlie behavioral differences between selectively bred groups of animals; (b) relations between physiological variables and individual differences in behavior, especially in the case of pathological deviations; (c) role of prenatal physiological factors in behavior development; (d) influence of early experience upon eventual behavioral characteristics; (e) cultural differences in child-rearing practices in relation to intellectual and emotional development; (f) mechanisms of somatopsychological relationships; and (g) psychological development of twins from infancy to maturity, together with observations of their social environment. Such approaches are extremely varied with regard to subjects employed, nature of psychological functions studied, and specific experimental procedures followed. But it is just such heterogeneity of methodology that is demanded by the wide diversity of ways in which he-

reditary and environmental factors interact in behavior development.

## REFERENCES

1. ANASTASI, ANNE. *Psychological testing.* New York: Macmillan, 1954.
2. ANASTASI, ANNE. *Differential psychology.* (3rd ed.) New York: Macmillan, 1958.
3. ANASTASI, ANNE, & CORDOVA, F. A. Some effects of bilingualism upon the intelligence test performance of Puerto Rican children in New York City. *J. educ. Psychol.,* 1953, **44,** 1–19.
4. ANASTASI, ANNE, & FOLEY, J. P., JR. A proposed reorientation in the heredity-environment controversy. *Psychol. Rev.,* 1948, **55,** 239–249.
5. ARSENIAN, S. Bilingualism in the postwar world. *Psychol. Bull.,* 1945, **42,** 65–86.
6. BARKER, R. G., WRIGHT, BEATRICE A., MYERSON, L., & GONICK, MOLLIE R. Adjustment to physical handicap and illness: A survey of the social psychology of physique and disability. *Soc. Sci. Res. Coun. Bull.,* 1953, No. 55 (Rev.).
7. BARKER, R. G., & WRIGHT, H. F. *Midwest and its children: The psychological ecology of an American town.* Evanston, Ill.: Row, Peterson, 1955.
8. BEACH, F. A., & JAYNES, J. Effects of early experience upon the behavior of animals. *Psychol. Bull.,* 1954, **51,** 239–263.
9. BRACKBILL, G. A. Studies of brain dysfunction in schizophrenia. *Psychol. Bull.,* 1956, **53,** 210–226.
10. DARCY, NATALIE T. A review of the literature on the effects of bilingualism upon the measurement of intelligence. *J. genet. Psychol.,* 1953, **82,** 21–57.
11. DAVIS, A., & HAVIGHURST, R. J. Social class and color differences in child rearing. *Amer. sociol. Rev.,* 1946, **11,** 698–710.
12. DOBZHANSKY, T. The genetic nature of differences among men. In S. Persons (Ed.), *Evolutionary thought in America.* New Haven: Yale Univer. Press, 1950. Pp. 86–155.
13. DOBZHANSKY, T. Heredity, environment, and evolution. *Science,* 1950, **111,** 161–166.
14. FORGUS, R. H. The effect of early perceptual learning on the behavioral organization of adult rats. *J. comp. physiol. Psychol.,* 1954, **47,** 331–336.

15. HALDANE, J. B. S. *Heredity and politics.* New York: Norton, 1938.
16. HARRELL, RUTH F., WOODYARD, ELLA, & GATES, A. I. *The effect of mothers' diets on the intelligence of the offspring.* New York: Bur. Publ., Teach. Coll., Columbia Univer., 1955.
17. HEBB, D. O. *The organization of behavior.* New York: Wiley, 1949.
18. HIRSCH, J., & TRYON, R. C. Mass screening and reliable individual measurement in the experimental behavior genetics of lower organisms. *Psychol. Bull.,* 1956, **53,** 402–410.
19. HORWITT, M. K. Fact and artifact in the biology of schizophrenia. *Science,* 1956, **124,** 429–430.
20. KALLMANN, F. J. *Heredity in health and mental disorder; Principles of psychiatric genetics in the light of comparative twin studies.* New York: Norton, 1953.
21. KING, J. A., & GURNEY, NANCY L. Effect of early social experience on adult aggressive behavior in C57BL10 mice. *J. comp. physiol. Psychol.,* 1954, **47,** 326–330.
22. LOEVINGER, JANE. On the proportional contributions of differences in nature and in nurture to differences in intelligence. *Psychol. Bull.,* 1943, **40,** 725–756.
23. LORENZ, K. Der Kumpan in der Umwelt des Vogels. Der Artgenosse als auslösendes Moment sozialer Verhaltungsweisen. *J. Orn., Lpz.,* 1935, **83,** 137–213; 289–413.
24. LUCHINS, A. S., & FORGUS, R. H. The effect of differential postweaning environment on the rigidity of an animal's behavior. *J. genet. Psychol.,* 1955, **86,** 51–58.
25. MELZACK, R. The genesis of emotional behavior: An experimental study of the dog. *J. comp. physiol. Psychol.,* 1954, **47,** 166–168.
26. MILNER, ESTHER A. A study of the relationships between reading readiness in grade one school children and patterns of parent-child interaction. *Child Develpm.,* 1951, **22,** 95–112.
27. PASAMANICK, B., KNOBLOCH, HILDA, & LILIENFELD, A. M. Socioeconomic status and some precursors of neuropsychiatric disorder. *Amer. J. Orthopsychiat.,* 1956, **26,** 594–601.
28. SCHWESINGER, GLADYS C. *Heredity and environment.* New York: Macmillan, 1933.

29. Scott, J. P., & Charles, Margaret S. Some problems of heredity and social behavior. *J. gen. Psychol.*, 1953, **48**, 209–230.

30. Scott, J. P., & Fuller, J. L. Research on genetics and social behavior at the Roscoe B. Jackson Memorial Laboratory, 1946–1951—A progress report. *J. Hered.*, 1951, **42**, 191–197.

31. Searle, L. V. The organization of hereditary maze-brightness and maze-dullness. *Genet. Psychol. Monogr.*, 1949, **39**, 279–325.

32. Snyder, L. H. The genetic approach to human individuality. *Sci. Mon., N. Y.*, 1949, **68**, 165–171.

33. Snyder, L. H., & David, P. R. *The principles of heredity*. (5th ed.) Boston: Heath, 1957.

34. Spoerl, Dorothy T. Bilinguality and emotional adjustment. *J. abnorm. soc. Psychol.*, 1943, **38**, 37–57.

35. Thompson, W. R., & Melzack, R. Early environment. *Sci. Amer.*, 1956, **194** (1), 38–42.

36. Tryon, R. C. Genetic differences in maze-learning ability in rats. *Yearb. nat. Soc. Stud. Educ.*, 1940, **39**, Part I, 111–119.

37. Whiting, J. W. M., et al. *Field guide for a study of socialization in five societies*. Cambridge, Mass.: Harvard Univer., 1954 (mimeo.).

38. Whiting, J. W. M., & Child, I. L. *Child training and personality: A cross-cultural study*. New Haven: Yale Univer. Press, 1953.

39. Williams, Judith R., & Scott, R. B. Growth and development of Negro infants: IV. Motor development and its relationship to child rearing practices in two groups of Negro infants. *Child Develpm.*, 1953, **24**, 103–121.

40. Woodworth, R. S. Heredity and environment: A critical survey of recently published material on twins and foster children. *Soc. Sci. Res. Coun. Bull.*, 1941, No. 47.

# HOW ARE INTERTRIAL "AVOIDANCE" RESPONSES REINFORCED?

O. H. MOWRER AND J. D. KEEHN [1]

*University of Illinois*

Beginning with a rather incidental observation reported by May (14) in 1948, a number of investigators (6, 22, 25) have shown that in a conditioning set-up, involving a noxious unconditioned stimulus, the incidence of intertrial (spontaneous) responses will increase markedly if such responses are allowed to function instrumentally, i.e., to "avoid," in the sense of postponing, the next "trial." This effect is not obtainable if the unconditioned stimulus is preceded by a warning signal (27), but does occur regardless of whether the instrumental intertrial response is the same as or different from the response produced by the unconditioned stimulus (13; cf. 18). The increased incidence of intertrial responses under the conditions described is in contrast to the previously reported observation that intertrial responses occur progressively *less* frequently in a conditioning situation wherein they cannot function instrumentally, i.e., where the intertrial interval is fixed or at least variable only by the experimenter (7, 19).

The discovery that an instrumental procedure of the kind described will increase the incidence of intertrial responses followed logically enough from the finding of Hunter (11), Brogden, Lipman, and Culler (3), and others that a response made to a warning (conditioned) stimulus is more readily and more reliably fixated if, when such a response occurs, the unconditioned stimulus is omitted, rather than being paired willynilly (as in "classical" condition-

ing) with the conditioned stimulus. And from this and related work it was soon evident that so-called avoidance learning is not a matter of simple (stimulus-substitution) conditioning, but that it involves two distinct stages or "levels": when a signal is paired with some painful stimulus, what an organism learns first is to *be afraid;* and then it learns, perhaps by a quite different process, what to *do* about its fear (16, 17). Yet, with respect to the learning of instrumental (avoidance) intertrial responses, there has been a tendency to adhere to the older one-stage stimulus-substitution type of interpretation. The purpose of the present paper is to show, on both empirical and logical grounds, how much more satisfactory is a two-factor type of explanation.

## PREVIOUS RESEARCH AND THEORY

In the winter of 1946, Hart Westbrook made what seems to have been the first attempt to set up an experimental situation in which intertrial responses might function instrumentally in the manner already indicated. However, since negative results were obtained (probably because training was not continued long enough; see later section on Interval-Avoidance Learning), the details of this study need not be described here. Then, as already indicated, two years later May reported the occurrence of intertrial ·response learning in a situation designed to investigate quite a different problem. This decidedly incidental, but striking, finding was reported thus:

In the above experiments the swinging door over the barrier [separating the two sides of

a shuttle box] was never locked. The rats could cross and recross at will. An indefinite number of crossings was possible during the training and testing trials. Sixty to 90 sec. were allowed between the last spontaneous crossing and the next training or testing trial. Thus, on the training trials an animal could avoid shock for a considerable time by crossing and recrossing at intervals of less than 60 sec. Ten of the 12 experimental animals (Group A—door closed) and 11 of the controls learned this form of adaptive behavior and used it fairly regularly both on the training and on the test trials. Very few of the animals in B-Group (door open), on the other hand, learned it (14, p. 71).

In the summary of the study cited, May does not allude to this finding and elsewhere in his paper makes no attempt to interpret or analyze it. However, in a study by Bugelski and Coyer (6), briefly reported in 1950, the problem was attacked quite directly and with a theoretical issue in mind, implied by the title, "Temporal Conditioning vs. Anxiety Reduction in Avoidance Learning." A few years later, Bugelski described the procedure employed thus:

Every 15 seconds a shock was applied in sufficient strength to force a rat over a hurdle. With repeated trials, the animals came to give almost perfect performance, jumping the hurdle in the general interval of 10 to 15 seconds, thus avoiding shock (5, p. 68).

In accompanying graphs Bugelski shows that, although rats learned to "shuttle" more readily under the conditions indicated (i.e., when the shock-shock interval was 15 seconds), another group of animals also solved the problem with a shock-shock interval of 60 seconds. Say Bugelski and Coyer of their study:

The results support Hull's account of avoidance behavior (*Principles*) as antedating responses developing from escape reactions. Anxiety reduction and expectancy explanations need not be invoked for the responses studied (6, p. 265).

By this, these writers apparently mean that shuttling, when it occurs between

trials, is just the conditioned (antedating) form of the escape reaction elicited by the shock which now, however, is produced by the experimental situation alone (without shock). Such an interpretation is, in a sense, classically Pavlovian and, as these writers point out, does not require that fear be posited as an intervening variable or that fear reduction be a form of reinforcement. Reasons for doubting the adequacy of this view will be given presently.

Then, in 1953, Sidman, using a different type of response and training procedure, also reported positive results:

White rats were the experimental organisms, with lever pressing selected as the avoidance response. Shocks of a fixed 0.2-sec. duration were given to the animal through a grid floor at regular [commonly 5-sec.] intervals unless the lever was depressed. Each lever depression reset the timer controlling the shock, thus delaying its appearance [by 20 seconds or so]. . . . Only the initial downward press on the lever reset the timer; holding the lever down had no effect upon the occurrence of the shock.

With no other contingencies between avoidance behavior and exteroceptive stimulation involved, approximately 50 animals have been successfully conditioned. . . . A striking characteristic of the initial curves is the abruptness with which the rate increases from its initial near-zero level. . . . With continued training the rate remains relatively stable not only within but also between sessions. Rates as high as 17 responses/min have been maintained by some animals during sessions totaling over 24 hours, with variations no greater than 0.1 responses/min appearing between the average rate for each session (25, pp. 157–158).

Sidman was inclined to interpret this striking effect as follows:

The behavior generated by this procedure can be explained by a model which holds that avoidance responding increases in rate at the expense of other behavior that is depressed by shock. An equivalent statement, in reinforcement terms, is that the avoidance response is strengthened when it terminates incompatible behavior that has been paired with shock (8, 23) [i.e., interval responses, other than the "correct" one, which do not postpone shock, get punished and thus tend to arouse fear]. Several lines of evidence indicate that the

avoidance rate is not simply some form of temporal conditioning in which the responses are triggered-off by the passage of a time interval (**25**, p. 158).

Because of the brevity of the paper just cited, the precise nature of Sidman's theoretical position was not entirely clear, so we turn to the paper by Schoenfeld which Sidman cites. This paper is long and intricate and cannot be summarized here; but the following sentences, from the summary, are perhaps the most relevant ones for present purposes:

> Following several objections raised against [the] anxiety-reduction hypothesis, an alternative explanation of avoidance is offered in which the response is conceived to be always a stimulus-terminating one. . . . It is argued that the avoidance response terminates [aversive] stimulus compounds in which proprioceptive and tactile stimuli are important components. . . . If this formulation is correct, avoidance conditioning proves to be a form of escape training, and its avoidance function is incidental to its stimulus-termination function (**23**, p. 97).

That the Schoenfeld position is, however, open to varying interpretations is indicated by the following comment by Kamin:

> The usual assumption has been that spontaneous responses result from conditioning to generalized apparatus cues (cf. **9**, p. 76; **23**, p. 89) (**12**, p. 71).[2]

While the theoretical picture was thus somewhat confused as of 1954, it is fair to say that the prevailing tendency was to explain the reinforcement of instru-

---

[2] Solomon and Brush (**33**, p. 266) allude to a study of intertrial avoidance learning reported (as a personal communication) by Milner in 1955. In only one respect (which will be mentioned later) does this study extend what had been learned from earlier experiments. Solomon and Brush also give a comprehensive review of research on avoidance learning where the CS-US interval is systematically varied (with different subjects). However, this line of work, while germane, is not directly relevant to the present discussion and will not be specifically cited.

mental intertrial responses as *not* involving the reduction of a temporally conditioned fear (or anxiety), and perhaps not involving the reduction of fear at all.

## EXPERIMENTAL DISPROOF OF CLASSICAL CONDITIONING INTERPRETATION

The fear-reduction hypothesis has been so successful in accounting for avoidance learning in general, and in resolving paradoxes generated by other theories, that there is strong presumption that it can also explain the learning of instrumental intertrial responses. But at once there is at least a superficial difficulty. In "ordinary" avoidance learning, i.e., where there is a definite and specific warning signal, it is simple and reasonable to suppose (*a*) that fear becomes conditioned to this signal and (*b*) that any response which turns the signal off (and averts the unconditioned stimulus) will be reinforced by fear reduction. But, in the case of instrumental intertrial responses, there is no such specific warning signal, so that the fear-reduction hypothesis, if applicable at all, must apply in some rather subtle and special way. In later sections, this inference will be considered in detail; but first a simple experiment will be reported which clearly shows the inadequacy of the classical conditioning type of explanation.

Kamin has already been quoted as saying: "The usual assumption has been that spontaneous responses result from conditioning to generalized apparatus cues." If the classical conception of conditioning were valid, then, every time a subject experienced electric shock in a given experimental situation and made a specific response thereto, there would be an increased tendency for the subject to make that response to the stimulus compound which immediately precedes the shock. Since it is the experimental situation as a whole, rather than a more specific stimulus, which here

precedes shock and thus becomes, in effect, the conditioned stimulus, the subject should soon start making the response in question to that situation "spontaneously," i.e., without a specific cue or signal.

On the other hand, if one accepts the two-factor version of avoidance learning, one must assume that, in order for intertrial responses to become more frequent, they must be reinforced by the occurrence of fear reduction. In the one case, mere contiguity of stimulation (situation + shock) would provide the conditions necessary for the increased occurrence of intertrial responses; whereas, in the other case, something more would be required: these responses would have to be in some sense functional, instrumental, problem solving. In the experiment to be described shortly, these two hypotheses are put to the test; but, before describing that experiment, it will be necessary to say a word about apparatus and procedure.

In the instrumental intertrial learning experiments already alluded to, Westbrook and Schaefer used a wheel-turning response, May and Bugelski and Coyer used shuttling, and Sidman used bar pressing. In casting about for an avoidance response that might be more "natural" for rats than either bar pressing or wheel turning and also free of the conflict that necessarily exists (at least in the beginning) in a shuttling set-up, one of the present writers (J. D. K.) tried out a revolving cage similar to that employed in the experiment of Brogden, Lipman, and Culler (3) and found it highly satisfactory. With the response now merely a short *run* (which revolves the cage but otherwise gets the subject nowhere), rats show excellent interval-avoidance learning; and with a few refinements the apparatus has proved itself to be reliable, flexible, and well adapted to the investigation of a wide range of related problems. The refine-

ments are these: (a) a simple, silent "lock" which allows the cage to move easily in one direction but not in the other; (b) a device which quietly resets an interval timer whenever the cage is turned (in the "correct" direction) by a minimum of two or three inches; and (c) articulation of the revolving cage with a cumulative graphic recorder (described elsewhere, 15). Thus, most of the procedure and all the recording are made automatic and put beyond the influence of experimenter produced variables.

Figure 1 shows the performance typical of an albino rat (a 140-day-old male) in the apparatus here described, under the following conditions. For one hour on three successive days this rat had been put into the revolvable cage, with no shock, and allowed to habituate. What little spontaneous running occurred in the beginning soon disappeared, so that on the third day the rat, in the course of an hour, turned the cage less than two revolutions, as can be seen by the virtually horizontal "baseline" in the lower left-hand corner of Fig. 1. Then, after the three daily sessions of habituation, the rat was subjected to daily training sessions, again of one-hour duration each, under the following conditions: a light electric shock from the grillwork which constitutes the "floor" of the revolvable cage came on every 20 seconds and remained on until the rat ran and revolved the cage far enough (roughly two inches) to turn off the shock, *unless* such a run occurred *between* shocks, in which event the timing mechanism reset and started another 20-second cycle, without shock. As will be seen in Fig. 1, this subject, during the first experimental session, received 37 shocks (indicated by the lateral spurs on the cumulative response line). Since a total of 180 shocks (3 to the minute for 60 minutes) would have been received during the experimental
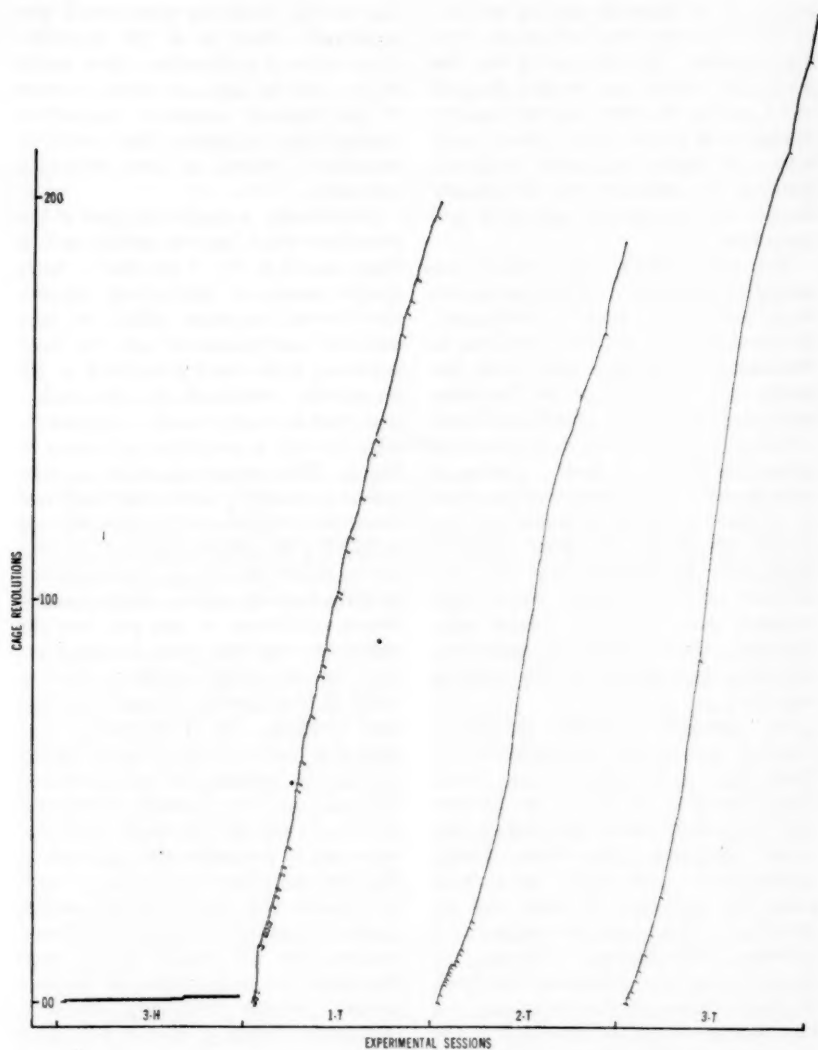
Fig. 1. Cumulative records of an albino rat's performance in a revolvable cage when between-trial ("spontaneous") runs postpone the next shock by 20 seconds. An activity baseline, representing the amount of running by the subject during the third of three daily habituation sessions of one hour each, under conditions of no shock, is shown in Graph 3-H. Graphs 1-T, 2-T, and 3-T represent the amount of running during hourly sessions on each of three successive "training" days. Shocks were applied only when the subject failed, for 20 seconds, to revolve the cage by the prescribed amount (two or three inches). Shocks are indicated by the lateral spurs on the underside of the graphs.

session if no interval running had occurred, it is clear that 143 shocks were thus avoided. On the second day the number of shocks not avoided dropped to 15, and on the third day this number dropped still further to 11. Here, manifestly, is highly successful avoidance learning, in a situation that is basically simple both as regards apparatus and procedure.

But such results, taken alone, are somewhat ambiguous. There are several ways in which they might be interpreted. In the absence of any other evidence to the contrary, one might infer, with Bugelski and Coyer, that the between-shock runs are simply conditioned (antedating) versions of the reaction of the subject to the shock itself. Lacking a more specific conditioned stimulus, such as a light or buzzer to herald the approach of shock, the *total situation* might here be thought of as the conditioned stimulus, which, being continuously present between shock presentations, elicits numerous, somewhat irregular occurrences of the running response.

Or, consider another possibility. There is an obvious resemblance between the curves shown in Fig. 1 and those commonly reported, by Skinner (31) and others, where the subject, motivated by hunger, gets reinforced only intermittently (with food) for pushing a bar but continues to make this response at a fairly rapid and regular rate between reinforcements. Unlike the hunger, which is continuous, the shock in this situation is discontinuous; but fear is presumably present much of the time while the subject is in the revolvable cage. So one might think of the response to shock-plus-fear as generalizing to fear alone and thus occurring many times on the basis of the sheer habit strength (or reflex reserve) generated solely by the reinforcement provided by escape from the shock-plus-

fear on the occasions when shock was presented. Here, as in the Bugelski-Coyer type of explanation, there would be no need to posit any reinforcement of the interval responses themselves, through fear reduction; they could be thought of instead as pure extinction responses.

Fortunately, a simple variation of the procedure which leads to results such as those shown in Fig. 1 provides a fairly crucial means of determining whether the interval responses which are here exhibited are reinforced only on those occasions when shock is received *or* are secondarily reinforced, by fear reduction, each time they occur. Typical results for such a procedure are shown in Fig. 2. Here an animal, which may be called C (control) and which had had three daily sessions of habituation just as had E (the animal used to obtain the curves shown in Fig. 1), was subjected to daily training sessions under the following conditions: it was put into the revolvable cage and given shocks at exactly the same points in time as those at which E had failed to "avoid" and had been shocked. In other words, since Animal E had received 37 shocks on the first day of training, Animal C received the same number of shocks distributed in time in exactly the same way; the same sort of procedure was repeated on the two remaining days. As a result of receiving the first 10 or 12 shocks, Animal C actually did somewhat more running than did Animal E, but soon thereafter the performance of the two animals became strikingly different. During the course of Session 1-T (first training day), Animal C, although receiving the same number (and approximate duration) of shocks, turned the cage only 94 revolutions as opposed to 198 revolutions for Animal E; and, in Session 3-T, the discrepancy had become much wider still: only 10 revolutions for Animal C as opposed to 249
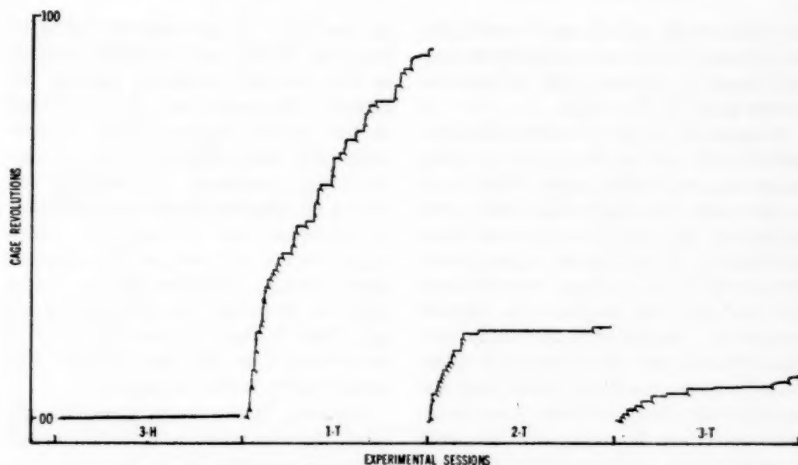
FIG. 2. Cumulative records of an albino rat's performance in a revolvable cage when a run (of two or three inches) turns off the shock (received from the cage floor) but, when occurring spontaneously, does not postpone the next shock. The four graphs are labeled as in Fig. 1. The shocks, equal in intensity and number, are distributed in time just as were those received by the subject whose performance is shown in Fig. 1. This type of procedure provides a control for certain hypotheses, cited in the text, which have been advanced to account for the type of interval-response avoidance learning shown in Fig. 1.

revolutions for Animal E. As Curve 3-T shows, when the shock came on, Animal C would still run far enough to turn the shock off (or perhaps a little further), but there was virtually no interval running. This animal had learned that the shocks were not avoidable, that they came regardless of what was done in the intervals between shocks; so, there being no gain from such behavior, the animal, sensibly enough, abandoned it. This outcome, substantiated by results obtained with three other pairs of animals, supports the hypothesis that interval responses, if they are to be perpetuated, must be reinforced in some manner other than by the reinforcement provided by the mere pairing of situation and shock.

INTERVAL AVOIDANCE LEARNING FURTHER ANALYZED AND INTERPRETED

From the empirical results just reported, it is evident that intertrial avoidance learning cannot be adequately explained in terms of classical conditioning. Some more intricate interpretation is obviously called for; the one which most readily suggests itself involves the now well authenticated concept of *stimulus trace* (1, 4). In an experimental situation of the kind under discussion, each response, whether forced (by shock) or spontaneous, presumably sets up a neural reverberation ("immediate memory") which decays to zero (or at least to asymptote) in roughly 30 seconds. If, therefore, electric shock be delivered to the subject whenever, let us say, the 20-second point on the stimulus trace is reached, one would expect that fear would become conditioned to this point on the trace and, in generalizing forward, would motivate the subject to make the response in question during the intertrial (20-second) interval, thus averting the impending shock. And, since each new occurrence of the

response would, so to say, "reset" the trace, there would be a reduction in fear and hence a reinforcement of the response itself (cf. 7, 20).[3]

Reasonable as it is a priori, this hypothesis did not, at first, seem to have much empirical justification. The Westbrook study was undertaken with this hypothesis explicitly in mind; and when preliminary work failed to reveal the expected effect, i.e., showed no tendency for the intertrial response to become temporally conditioned, the study was discontinued and the hypothesis questioned. May, as already noted, was apparently the first to obtain good intertrial avoidance learning, but he made no attempt to account for it theoretically. Bugelski and Coyer, to be sure, reported (6) that, with repeated trials, the animals came to give almost perfect performance, jumping the hurdle in the general interval of 10 to 15 seconds, thus avoiding shock; but these writers advanced the now untenable antedating hypothesis (of Pavlov and Hull). Sidman, finding no evidence of temporal conditioning in his own early researches (see Fig. 3), adopted an explanation, derived from Schoenfeld, to the effect that the "correct" interval response is learned because other intertrial responses

[3] Imagine an experimental situation in which a relatively loud tone came on immediately after the subject had responded in a specified way and then gradually softened—or a patch of light which, upon occurrence of the response, became suddenly large and then gradually diminished in size. Suppose also that the subject never received shock when the stimulus was large (or loud) but was shocked when it diminished beyond a certain point. It would be in no way surprising if the subject, under these conditions, learned to use such a stimulus as the basis of successful avoidance behavior. Our conjecture is that, in interval-avoidance learning of the kind under discussion, something comparable is involved, the principal difference being that the changing cue, instead of being an external stimulus, is a stimulus *trace* produced by the last preceding instrumental (overt) response.

get punished (in the sense of failing to postpone shock) and therefore, relative to the "correct" response, become inhibited. Moreover, since the occurrence of the correct response stops or prevents the occurrence of other (fear-producing) responses, this response will have a tendency to reduce fear and thus be reinforced—an interpretation which makes no use whatever of the stimulus-trace concept. (Sidman did not, to be sure, use precisely the above terminology; but, in light of considerations to be advanced in the next section, this paraphrasing seems legitimate.)

Schaefer, in the recent study already cited, likewise failed to find evidence of temporal conditioning in his subjects—in fact, their intertrial responses tended decidedly to pile up (as in Fig. 3) toward the beginning of the intertrial interval rather than toward the end thereof. And on the basis of this finding, Schaefer advanced an explanation of intertrial avoidance learning which, while in effect something like the one put forward by Sidman, is at least differently phrased. According to Schaefer's conjecture, what the subject in an experiment of this kind does is to *discriminate* between two conditions: situation with correct response being made and situation with this response *not* being made. And, since the rat is never shocked while thus responding and *is* shocked when responding differently or not at all, then the rat, feeling safe while responding correctly and fearful at all other times, is differentially reinforced on each occurrence of the correct response. In other words, by making a particular response, the animal, so to say, converts the situation from a dangerous one to a safe one and is thereby rewarded. Here the dilemma mentioned earlier—that there is no explicit danger signal in a situation of this kind whose termination can provide the specific occasion for fear reduction and reinforce-

RAT H-28
Session 3

FREQUENCIES

300

250

200
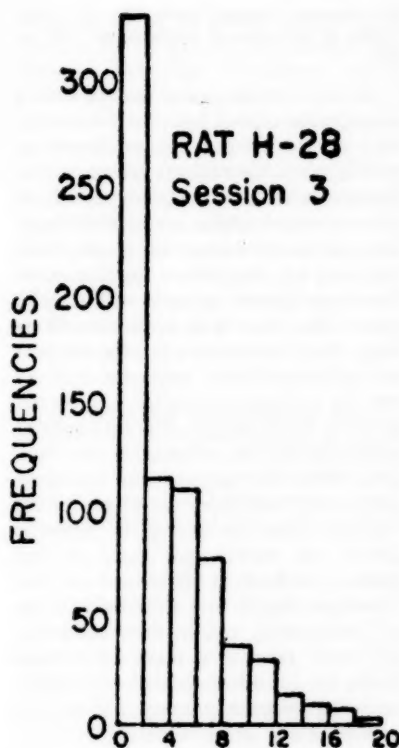
150

100

50

0

0    4    8    12    16    20

Fig. 3. Distribution of intertrial responses made by a rat during Session 3 of an experiment reported by Sidman (26). The abscissa represents seconds of time elapsing between the preceding response and the next one. When the subject delayed by as much as 20 seconds between responses, it then received an electric shock of brief (0.2 sec.) duration. As will be noted, most responses occurred shortly after the preceding response, rather than piling up in the true "danger zone," i.e., during the latter part of the intertrial period (see text).

ment of the avoidance response—is resolved by making the situation as a whole, *without correct responding,* the "danger signal," which is then "turned off" whenever and while the animal makes the requisite response.

This is an eminently satisfactory theory (and is stated more simply and

plausibly than Sidman's somewhat equivalent one); but it cannot, alone, account for all the now known facts. In 1954, in a paper entitled "The Temporal Distribution of Avoidance Responses," Sidman reported that "after continued training" there is a tendency for a "well developed time-discrimination" to appear (26, p. 401). In two recent papers, Boren and Sidman (2, 29) show that, if training is continued long enough (to or beyond approximately 100 hours), then interval-avoidance responses do indeed take on the expected temporal distribution, i.e., they tend to pile up near the remote end of the intertrial interval, rather than toward the beginning thereof (see Solomon and Brush's account of Milner's work, 33, p. 266). Hence, even though the Sidman-Schaefer type of explanation be valid for the results first obtained in interval-response avoidance learning, such an explanation must obviously be supplemented if one is also to account for the opposite type of temporal distribution of responses obtained after more extended training. One possibility, therefore, is that, in experimentation of the kind under discussion, the *first* discrimination that is developed is a comparatively gross one, i.e, between experimental situations with and without the correct response being in progress, and that a *second* stage of discrimination later develops in which the subject discovers that situation-without-correct-responding is dangerous *only* after an interval of time has elapsed.

There is, however, also the possibility that the stimulus-trace type of explanation can account for *all* the facts. Let us assume that fear is most strongly conditioned to a point on the stimulus trace 20 seconds removed from the event that sets up the trace (i.e., the preceding correct response, either forced or spontaneous). Let us further assume, as we reasonably may, that fear (at least

in the beginning) *generalizes far forward* on the trace. If, then, as already posited, occurrence of the correct response during the intertrial period sets up a new trace and reduces fear, one might expect this response to occur just as soon as the subject *starts* being afraid. In other words, there is nothing in the nature of the situation which would require that the subject wait until *maximally* afraid before reacting defensively. Therefore, the temporal distribution of the incidence of the correct overt response would not be expected necessarily to provide a faithful picture of the temporal pattern of fear intensity. Occurrence of the correct overt intertrial response might very well pile up, not where the fear would be maximally intense, but rather where it is first experienced.

Eventually, of course, one would expect discrimination to develop and to offset this original overgeneralization, so that fear would be first experienced during the intertrial period, not far forward (due to generalization along the trace), but instead toward the latter part of the period where the objective danger really is.

In a study published in 1956 under the title "Time Discrimination and Behavioral Interaction in a Free Operant Situation," Sidman has shown that it is a simple matter to get rats to respond with considerable accuracy to a 20-second temporal interval when they are thirst-motivated and water-reinforced for pressing one bar, at the appropriate interval, after having pressed (on cue) another bar. Thus, as Sidman observes:

The procedure is an operant analogue of Pavlovian trace conditioning, with the qualification that the "trace" is initiated not by an exteroceptive stimulus but by the organism's own behavior. . . . The response probability increases with the passage of time, not because of the dissipation of inhibition [as Hull (10) is credited with holding], but rather because the situation becomes more like that prevailing at the time of reinforcement (28, p. 469).

At first it may appear strange that a free-operant type of behavior (bar pressing) can be temporally conditioned so readily when the subjects are thirst-motivated; whereas, as already noted, in an avoidance-learning situation the same free operant for a long time occurs much too soon, i.e., long before the time when the unconditioned stimulus will actually occur. But there is an instructive difference: in the situation when the subjects are thirst-motivated, responses that occur too soon go unrewarded, i.e., do not produce water and are thus extinguished relatively rapidly; whereas, in the situation when the subjects are *fear*-motivated, each and every occurrence of the "correct" response during the intertrial period, no matter how early in that period, produces a decrement in fear (however slight) and thus tends to be self-reinforcing, rather than self-extinguishing. Here, as in many other situations, we see the importance of differentiating between primary and secondary drive and drive reduction.

## Two Different Conceptions of "Aversive" Stimulation

In earlier sections of this paper, we have seen how numerous and determined the attempts have been to account for various forms of avoidance behavior without recourse to the concept of fear as an intervening variable. There is, of course, every justification in science for trying to push the principle of parsimony as far as possible. But it must be remembered that sometimes a given type of parsimony proves to be quite impossible and has to be replaced by a more complex conceptual scheme. It is the present writers' belief that a simple one-step (S–R) conception of behavior is inadequate and that, at the very least, we must posit a two-step S–r:s–R model,

where S and R, as in the simpler S–R scheme, are "objective" stimulation and response, but where, in addition, r and s are intervening variables, of which *fear* (as both a reaction and as a drive) is a prime example.

The attempt of Sheffield to interpret the Brogden-Lipman-Culler experiment (previously cited) solely in terms of the contiguity (stimulus-substitution) principle and thus to eliminate the necessity of positing fear as an intervening variable illustrates the difficulties which such efforts commonly encounter. Says Sheffield:

The findings for avoidable shock offer no evidence for a strengthening effect of avoidance and follow the conventional result in Pavlovian experiments. Reinforcement by shock strengthened conditioned running; omission of shock led to extinction (24, pp. 174–175).

The point can be illustrated by referring back to Fig. 1. There it will be seen that, in Sessions 2-T and 3-T, in each of the two longest runs of interval-responding, there is, indeed, a tendency for this behavior to extinguish (as evidenced by the flattening of the cumulative response curve) and to have to be eventually reinforced by shock presentation. Here, it might seem, is clear evidence that shock avoidance is not reinforcing and that the response of running is indeed dependent upon the occasional occurrence of the shock. The difficulty, of course, is the one that always arises in such a situation as a result of speaking as if only *one* response were involved. The present writers assume that shock is, of course, occasionally necessary to reinforce the *fear* response, but that the *running* response is reinforced by fear reduction not by shock. It is our assumption that, whenever enough fear is present to motivate intertrial running, there will be an ensuing experience of fear reduction, with the result that this behavior is always rein-

forced. It is, rather, the underlying fear that extinguishes and has to be eventually reinforced by presentation of shock. Our assumption is that, if fear could be maintained at an appropriate intensity by some other means, the intertrial running would *never* extinguish. This assumption is supported by the recently reported finding of Sidman, Herrnstein, and Conrad that:

The effect of occasional free shocks, delivered independently of the monkeys' behavior, was to increase the rate of avoidance responding. . . . The effect of the free shock persisted over as long as 300 experimental hours during which time no other shocks were delivered (30, p. 557).

The authors' interpretation of this effect, in terms of Skinner's theory of superstitious behavior, seems to us contrived and nonparsimonious.

In an unpublished paper, Sidman, Herrnstein, and Boren try—but, we believe, unsuccessfully—to develop a conceptual scheme which will account for all forms of avoidance learning without recourse to the concept of fear as a motivating and reinforcing agent. In essence, the position is pristine behaviorism and stems directly, as the authors acknowledge, from the empirical researches and methodological stance of B. F. Skinner. In his 1953 book, *Science and Human Behavior*, Skinner states his position in this connection with stark simplicity. The chapter on "Emotion" begins with this sentence: "The 'emotions' are excellent examples of the fictional causes to which we commonly attribute behavior" (32, p. 160). Then the familiar, but highly debatable, assertion by William James is approvingly quoted to the effect

that we feel sorry because we cry, angry because we strike, afraid because we tremble, and not that we cry, strike, or tremble because we are sorry, angry, or fearful, as the case may be (32, p. 160).

And, later in the chapter, there is a section specifically headed "Emotions Are Not Causes."

Rarely are theoretical issues (in psychology) so beautifully explicit and "clean." The present writers, along with many others, take the position that the emotion of fear *is* causal; that a one-step (simple S–R) psychology will not work (17, 21); that, as a minimum, two causal steps are necessary to account for behavior; and that fear, in the case of so-called avoidance behavior, is an essential intermediate cause or determinant.

On the other hand, Skinner and those of like persuasion hold that this is not at all the case and that a "chaining interpretation," not greatly different from the earlier Watsonian notion of reflex chaining, is "capable of handling most of the data and of extension to other findings, while at the same time maintaining a degree of parsimony and internal consistency that we were unable to achieve by means of the alternative considerations" (unpublished manuscript by Sidman-Herrnstein-Boren).

This is not the time or place to enter into a detailed critique of the chaining interpretation, but one particularly questionable aspect thereof may be pertinently noted, namely, the tendency to speak of stimuli which have been paired with punishment as being aversive. This, we believe, involves a logically indefensible attempt to avoid the more likely possibility that such stimuli, after conditioning, become aversive because, and only in the sense that, they now elicit *fear* and that it is this latter state, and not the eliciting stimulus, which provides the aversive, motivating element in the situation (cf. Solomon and Brush, 33, p. 245 ff.).

## SUMMARY

It is now empirically established (*a*) that responses which occur between trials in a noxious conditioning situation occur progressively less frequently if trials occur ineluctably, i.e., at fixed intervals, and (*b*) that between-trial responses occur progressively *more* frequently when the next trial is delayed each time such a response is made. There is, of course, no mystery as to why intertrial (spontaneous) responses disappear in the former situation: they accomplish nothing and, if they happen to occur late in the intertrial interval, the ensuing noxious stimulus acts, in effect, as a punishment.

But there has been considerably less agreement as to how it is, under the second set of conditions mentioned, that intertrial responses are reinforced and perpetuated. A number of writers have taken the position that here the intertrial response (made to the experimental situation as a whole, rather than to a specific danger signal) is just the antedating (classically conditioned) form of the response made to the noxious, unconditioned stimulus on the trial. In the present paper various considerations are cited which indicate that this explanation is too simple and that a more complex explanation, involving the intervening advents of fear arousal and fear reduction, is necessary and generally sufficient.

Intertrial avoidance learning, as a laboratory phenomenon, deserves continued study, both because of its intrinsic theoretical interest and because of its rather patent relationship to compulsive (neurotically repetitive) behavior and to the psychology of *work*.

## REFERENCES

1. ANDERSON, O. D., & LIDDELL, H. S. Observations on experimental neurosis in sheep. *Arch. Neurol. Psychiat.*, 1935, 34, 330–354.

2. BOREN, J. J., & SIDMAN, M. Maintenance of avoidance behavior with intermittent shocks. *Canad. J. Psychol.*, 1957, 11, 185–192.

3. Brogden, W. J., Lipman, E. A., & Culler, E. The role of incentive in conditioning and extinction. *Amer. J. Psychol.*, 1938, **51**, 109–117.

4. Brown, J. S. A note on a temporal gradient of reinforcement. *J. exp. Psychol.*, 1939, **25**, 221–227.

5. Bugelski, B. R. *The psychology of learning.* New York: Holt, 1956.

6. Bugelski, B. R., & Coyer, R. A. Temporal conditioning vs. anxiety reduction in avoidance learning. *Amer. Psychologist*, 1950, **5**, 264–265 (Abstract).

7. Coppock, H., & Mowrer, O. H. "Spontaneous" responses as "rehearsal": A study of "overt thinking" in animals. *Amer. J. Psychol.*, 1947, **60**, 608–616.

8. Hefferline, R. F. An experimental study of avoidance. *Genet. Psychol. Monogr.*, 1950, **42**, 231–334.

9. Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

10. Hull, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.

11. Hunter, W. S. Conditioning and extinction in the rat. *Brit. J. Psychol.*, 1935, **26**, 135–148.

12. Kamin, L. J. Traumatic avoidance learning: The effects of CS-US interval with a trace-conditioning procedure. *J. comp. physiol. Psychol.*, 1954, **47**, 65–72.

13. Keehn, J. D. On the nonclassical nature of avoidance behavior. *Amer. J. Psychol.*, in press.

14. May, M. A. Experimentally acquired drives. *J. exp. Psychol.*, 1948, **38**, 66–77.

15. Mowrer, O. H. A cumulative graphic work-recorder. *J. exp. Psychol.*, 1943, **33**, 159–163.

16. Mowrer, O. H. On the dual nature of learning—a re-interpretation of "conditioning" and "problem solving." *Harvard educ. Rev.*, 1947, **17**, 102–148.

17. Mowrer, O. H. Two-factor learning theory reconsidered, with special reference to secondary reinforcement and the concept of habit. *Psychol. Rev.*, 1956, **63**, 114–128.

18. Mowrer, O. H., & Lamoreaux, R. R. Fear as an intervening variable in avoidance conditioning. *J. comp. Psychol.*, 1946, **39**, 29–50.

19. Mowrer, O. H., & Lamoreaux, R. R. Conditioning and conditionality (discrimination). *Psychol. Rev.*, 1951, **58**, 196–212.

20. Mowrer, O. H., & Viek, P. An experimental analogue of fear from a sense of helplessness. *J. abnorm. Soc. Psychol.*, 1948, **43**, 193–200.

21. Osgood, C. E. *Method and theory in experimental psychology.* New York: Oxford Univer. Press, 1953.

22. Schaefer, V. H. Avoidance conditioning in the absence of external stimulation: Some experimental and genetic parameters. Unpublished doctoral dissertation, Univer. of Illinois, 1957.

23. Schoenfeld, W. N. An experimental approach to anxiety, escape, and avoidance behavior. In P. H. Hock & J. Zubin (Eds.), *Anxiety.* New York: Grune & Stratton, 1950.

24. Sheffield, F. D. Avoidance training and the contiguity principle. *J. comp. physiol. Psychol.*, 1948, **41**, 165–177.

25. Sidman, M. Avoidance conditioning with brief shock and no exteroceptive warning signal. *Science*, 1953, **118**, 157–158.

26. Sidman, M. The temporal distribution of avoidance responses. *J. comp. physiol. Psychol.*, 1954, **47**, 399–402.

27. Sidman, M. Some properties of the warning stimulus in avoidance behavior. *J. comp. physiol. Psychol.*, 1955, **48**, 444–450.

28. Sidman, M. Time discrimination and behavioral interaction in a free operant situation. *J. comp. physiol. Psychol.*, 1956, **49**, 469–473.

29. Sidman, M., & Boren, J. J. Avoidance behavior maintained by shock-correlated variable response-shock intervals. *J. comp. physiol. Psychol.*, in press.

30. Sidman, M., Herrnstein, R. J., & Conrad, D. G. Maintenance of avoidance behavior by unavoidable shocks. *J. comp. physiol. Psychol.*, 1957, **50**, 553–557.

31. Skinner, B. F. *The behavior of organisms.* New York: Appleton-Century Crofts, 1938.

32. Skinner, B. F. *Science and human behavior.* New York: Macmillan, 1953.

33. Solomon, R. L., & Brush, Elinor S. Experimentally derived conceptions of anxiety and aversion. *Nebraska symposium on motivation.* Lincoln: Univer. Nebraska Press, 1956.

# THE DERIVATION OF SUBJECTIVE SCALES FROM JUST NOTICEABLE DIFFERENCES [1]

## R. DUNCAN LUCE

*Department of Social Relations, Harvard University*

## AND WARD EDWARDS

*Operator Laboratory, Air Force Personnel and Training Research Center*

The study of cumulated jnd scales began with Fechner; Fechner's law is such a scale. Psychophysicists have been deriving such scales and comparing them with scales derived in other ways, notably by fractionation, ever since, and a lot of controversy has resulted. The controversy is particularly hot at present because Stevens and Galanter (16) and Stevens (14) have assembled a lot of data which indicate that cumulated jnd scales do not agree with magnitude scales derived by other methods for intensity continua such as loudness, brightness, and pain.

Unfortunately, Fechner's procedure for cumulating jnds, which has been widely defended but not widely applied since his day, rests on an assumption which is inconsistent with one of his definitions. This means that cumulated jnd scales developed by his procedure are incorrect, and so comparisons between them and other kinds of scales are meaningless.

This paper begins by showing that Fechner's method contains internal contradictions for all but a few special cases, and that it cannot be rescued by minor changes. It goes on to derive a new and mathematically appropriate method for cumulating jnd's. This method turns out to be the simplest possible one: you can best cumulate jnd's simply by adding them on top of each other, like a stack of plates. Unfortunately, the detailed mathematical equivalent of this very simple operation is often fairly complicated. A simple but sometimes tedious graphic procedure, however, is readily available—and indeed has customarily been used by most scientists when developing cumulated jnd scales. This paper ends by discussing practical applications of this method, the relation it bears to scaling methods based on the law of comparative judgment, and the current controversy about scaling methods in psychophysics.

*The model of a sensation scale.* The psychophysical model of a sensation scale is a mathematical model; a sensation scale is an intervening variable. The rules by which sensation scales should be constructed are to some degree arbitrary, limited by logic, convenience, intuition, and best fit to data.

The model of a sensation scale goes as follows. Corresponding to many of the major subjective dimensions of change of sensory experience, there are primary physical dimensions of change (e.g., pitch and frequency, loudness and amplitude, etc.). Once parametric conditions for significant variables have been specified, we assume that a single-valued, monotonic, everywhere differentiable (smooth) function exists that relates the subjective dimension to its corresponding physical dimension. From here on, we shall use the words "dimension" and "continuum" interchangeably; we shall usually talk about a stimulus continuum and its corresponding sensory continuum.

That is the model, and it is very easy to state. The big difficulty comes when we try to decide how to fit data to it. All methods for doing this must introduce definitions and assumptions beyond those listed in the previous paragraph. These differ from one method to another.

The oldest sensory scaling method, Fechner's, is based upon a further condition that says that any jnd on a given sensory continuum is subjectively equivalent to any other jnd on that continuum. Whether this added condition is to be interpreted as merely a definition of the scale under consideration or as an assumption is a matter of opinion. Textbooks usually say Fechner "assumed" that all jnd's for a sensory dimension are equal to one another (1). It is not easy to know what he had in mind, but judging by his writings he probably did view it as an assumption having implications beyond scale construction. Since it is not directly observable and since its indirect consequences are highly debatable, others since Fechner have suggested that it might better be viewed as a definition.

It is our view that this is the more sensible position; it certainly is the one for us to take in this paper since our point is a logical one, not a substantive one. If equality of jnd's is taken as a definition, then it cannot be proved or disproved by any kind of empirical evidence. An experiment, for instance, that showed that a tone 20 jnd's loud is not half as loud (according to fractionation judgments) as a tone 40 jnd's loud would not have any relevance to what we may call Fechner's definition; it would only show that the kind of sensation scale implied by his definition does not agree with the kind implied by the definitions used in fractionation experiments. The issue, then, becomes what are the different scales useful for, and what is their relationship one to another. The latter part of this paper touches briefly on this problem.

The main purpose of the paper is to explore the consequences of Fechner's definition. Therefore, we must be certain of its meaning. It clearly does not mean, for instance, that all jnd's for loudness contain the same number of physical units. It is only on the sensory continuum, not on the stimulus continuum, that jnd's are defined as equal to one another. Furthermore, this definition holds only if all stimulus properties except those on the primary stimulus continuum remain constant. So there is no reason, for instance, to expect that the subjective size of a loudness jnd at 1,000 cycles per second (cps) should be the same as that of a loudness jnd at 4,000 cps. Of course, it would be pleasant if they were equal in size, but the model does not require it.

From here on we will talk about two kinds of jnd's. A sensation jnd is the magnitude of a jnd as measured in the units of the appropriate sensa-

tion continuum. By definition, all sensation jnd's for a given sensation continuum are equal to one another, given unchanged values of all stimulus properties except those on the primary stimulus continuum. A stimulus jnd is the magnitude of the change on the primary stimulus continuum, measured in appropriate physical units, which is just sufficient to produce a change of one sensation jnd upward at that point. ( A discussion of the essentially statistical nature of jnd's appears later in this paper.) In general, stimulus jnd's will have different sizes at different points on the primary stimulus continuum. The rest of this paper will not be intelligible unless you keep the distinction between these two kinds of jnd's in mind.

We have assumed that jnd's are measured upward on the stimulus continuum. They could also be measured downward, and the possibility exists that the two measurements might not agree. In fact, they are certain not to agree if the distance spanned is more than two jnd's, and if the size of the jnd at the end where measurement starts is used as the unit of measurement, since this means that the size of the measurement unit will be different depending on direction of measurement. However, such discrepancies might exist in the measurement of a single jnd; this, if it happened, would mean that jnd's are not suitable units of measurement unless direction is specified. We have therefore confined ourselves to upward jnd's.

Now we can say exactly what this paper is about. Given a function (obtained from experiment, theory, or both) relating stimulus to sensation jnd's for all points of the primary stimulus continuum, what may we infer about the sensory scale implied by that jnd function?

*Fechner's derivation of Fechner's law.* On October 22, 1850, Fechner (2) thought up the first (incorrect) answer to the question which ended the previous paragraph. Let us call any function that gives the size of a stimulus jnd at each point of the stimulus continuum a *Weber function* (corresponding to "a function relating stimulus to sensation jnd's" of the previous section), and any one-to-one function based on cumulated jnd's which relates the stimulus continuum to a sensory scale a *Fechner function* (corresponding to "a sensation scale" of the previous section). These definitions do *not* restrict our attention to those two special functions which have come to be known in psychophysics as Weber's law and Fechner's law! Fechner believed that the Fechner function corresponding to any Weber function could be expressed as the solution (integral) of a first order linear differential equation involving that Weber function. He applied this procedure to Weber's aw, which asserts that for a given stimulus continuum the size of the stimulus jnd, $\Delta x$, divided by the value of the stimulus at that point, $x$, is a constant ($\Delta x/x = k$). Let us examine his argument.

If Weber's law is true, then, since all sensation jnd's are equal by definition, there is a constant $A$ such that

$$\frac{\Delta u}{\Delta x} = \frac{A}{x} \qquad [1]$$

where $\Delta u$ denotes the size of the sensation jnd. The heart of Fechner's solution to his and our basic problem was to "rewrite" Equation 1 as the differential equation

$$\frac{du}{dx} = \frac{A}{x} \qquad [2]$$

How did Fechner make this step from differences (deltas) to differentials? He used what he called a "mathematical auxiliary principle," the essence of which is that what is true for differences as small as jnd's ought also to be true for all smaller differences and so true in the limit as they approach zero (differentials). If this argument were acceptable (which it is not), the rest would be simple. Equation 2, when integrated, yields the familiar logarithmic relationship between sensation and stimulus which is known as Fechner's law.

Fechner thought that his general procedure ought to be applicable to any Weber function, not just to Weber's law. It is not. Except for a few special cases like Weber's law, the definition of sensation jnd's as equal and the "mathematical auxiliary principle" are mutually contradictory. For example, consider the Weber function $\Delta x/x^2 = k$. Then, following Fechner's procedure, we should write:

$$\frac{\Delta u}{\Delta x} = \frac{A}{x^2} \quad \text{and so} \quad \frac{du}{dx} = \frac{A}{x^2}$$

Integrating, we get

$$u = B - \frac{A}{x}$$

Let us now check to see whether this new Fechner function satisfies the definition which says that sensation jnd's are equal to one another. If we are at point $x$ on the stimulus continuum, a stimulus jnd, according to the Weber function used in this example, is $kx^2$. The sensation increment corresponding to this change, the sensation jnd at this point, is therefore given by:

$$u(x + kx^2) - u(x)$$
$$= B - \frac{A}{(x + kx^2)} - B + \frac{A}{x}$$
$$= \frac{Ak}{1 + kx}$$

which is clearly not a constant for any value of the constant $A$ except zero.

This, although only one example, is typical in the sense that almost any example you could think of would show the same discrepancy. Only for a very few Weber functions—some pathological ones, Weber's law, and its generalization $\Delta x = kx + c$—does the "mathematical auxiliary principle" yield a Fechner function with equal jnd's. We will not take space to prove this formally, but a formal proof is available.

*The functional equation solution.* We have shown that Fechner's procedure involves a self-contradiction. We shall show later that it leads to wrong results in all important cases except Weber's law. Obviously the "mathematical auxiliary principle" is wrong and must go.

How, then, should we cumulate jnd's? The simplest, most obvious procedure (which has very often been used exactly because it is simplest and most obvious) is simply to add them up one at a time. If the first jnd on a primary stimulus continuum is 5 stimulus units, then two points on our cumulated jnd scale should be 0, 0 and 1, 5, where the first number is the scale value on the $y$ axis and the second number is the corresponding stimulus value on the $x$ axis. If we then find that the size of the stimulus jnd at 5 on the stimulus continuum is is 8, then the third point is 2, 13. If we find that the size of the stimulus jnd at 13 is 10, then the fourth point is 3, 23, and so on.

Fechner and some of his more modern imitators went way out of their way to avoid this simple and sensible procedure; in retrospect it is hard to decide why they did so. At any rate, the next two sections of this paper will develop a formal mathematical solution to Fechner's mathematical problem—a solution which turns

out to be the mathematical equivalent of the simple graphical or arithmetic technique discussed in the previous paragraph. The mathematical problem centers about how to fill in the curve between the discrete points arrived at by the graphical method.

What mathematical tools can we use to replace Fechner's "mathematical auxiliary principle"? Equation 1, and the corresponding ones based on other Weber functions, can be solved directly without any mathematical auxiliary principles or other further assumptions. They are examples of what mathematicians call functional equations. The papers on which most of our discussion is based (5, 6) were published in the 1880s, twenty years after Fechner first published his work.

The kind of functional equation implied by the definition of equality of sensation jnd's is soluble for a very wide class of Weber functions. Unfortunately, there is an infinity of inherently different solutions to each of these equations. However, further consideration of what we mean by a sensation scale will lead us to properties which we usually take for granted and which are enough to narrow the solutions down to just one interval scale, unique except for its zero point and unit of measurement. It is interesting that in the case of the linear generalization of Weber's law, and in that case only, the functional-equation solution is the same as that obtained by Fechner's auxiliary principle; for all other Weber functions the two solutions are different.

First, we will state the general mathematical problem and its solution. Let $x$, $x \geq 0$, denote a typical value of the stimulus continuum, and let $u$ denote the (unknown) Fechner function. Let $g$ be the (given) Weber function; i.e., a stimulus magnitude

$y$, $y \geq x$, is detected as larger (in a statistical sense) than $x$ if $y > x + g$ $(x)$, whereas it is not discriminated as different from $x$ if $x \leq y \leq x + g(x)$. We write $x + g(x) = f(x)$. By definition, a sensory jnd at the sensation $u(x)$ is given by the increment[2]

$$u[f(x)] - u(x)$$

(In the usual "delta" notation, $g(x) = \Delta x$ and $u[f(x)] - u(x) = \Delta u$.) The condition that sensation jnd's be equal simply means that all sensation jnd's are a constant, which we may take to be 1 for convenience, since an arbitrary change of unit does not matter. Thus, we have our major mathematical problem:

Find those real-valued differentiable functions $u$, defined for all $x \geq 0$, such that $u[f(x)] - u(x) = 1$, for all $x \geq 0$.

Note that we have said *those* functions, not *that* function, for there may be more than one such function. This uniqueness question has not traditionally been raised, for so long as the problem was formulated in terms of linear differential equations, the uniqueness theorems of that branch of mathematics insured only one solution. In the realm of functional equations, we have no such assurances.

It is very lucky that the functional equation which has arisen in this problem is one of the more famous in the

[2] Throughout this paper we shall have to use functions of functions. In general, if $v$ and $w$ are two real-valued functions of a real variable $x$, $v[w(x)]$ denotes the number obtained by calculating $y = w(x)$ and then finding $v(y)$. Clearly, the order of writing $v$ and $w$ is material, for $v[w(x)]$ does not generally equal $w[v(x)]$. Consider, for example, $v(x) = ax$, where $a \neq 1$, and $w(x) = x^2$. Then, $v[w(x)] = v(x^2) = ax^2$, whereas $w[v(x)] = w(ax) = a^2x^2$.

literature; it is called Abel's equation.[3] The principal results we shall need concerning this equation were presented by Koenigs (5, 6) in 1884 and 1885.[4] First, we will present his uniqueness results, which illustrate the method of attack and lead up to the general solution. Suppose that $u_0(x)$ is a solution to Abel's equation, and suppose $p(x)$ is an arbitrary periodic function with period 1—in other words, any function satisfying

$$p(x + 1) = p(x)$$

$K \sin 2\pi x$ is periodic with period 1, and so is an example of a function $p(x)$. It is easily to show that the function $u_p(x) = u_0(x) + p[u_0(x)]$ is also a solution to Abel's equation:

$$
\begin{aligned}
u_p[f(x)] &= u_0[f(x)] + p\{u_0[f(x)]\} \\
&= 1 + u_0(x) + p[1 + u_0(x)] \\
&= 1 + u_0(x) + p[u_0(x)] \\
&= 1 + u_p(x)
\end{aligned}
$$

Furthermore, it can be shown that if $u$ and $u^*$ are two solutions to Abel's equation, then there exists a periodic function $p$ with period 1 such that

$$u(x) = u^*(x) + p[u^*(x)]$$

Thus, if we have any solution $u_0$ to our problem and if we choose $p$ to be a differentiable periodic function with period 1, then $u_p = u_0 + p(u_0)$ is also differentiable and solves the problem.

In the case of Weber's law, we have $f(x) = kx$, $k > 1$, and the differentiable function $u_0(x) = \dfrac{\log x}{\log k}$ is

[3] Sometimes this equation is spoken of as the Abel-Schroder equation, but more often Abel's name is attached to this equation and Schroder's name to the equation $v[f(x)] = cv(x)$, which arises from Abel's equation through the substitution $v = c^u$.

[4] We are indebted to Richard Bellman of the RAND Corportion for directing us to the literature on the Abel equation.

easily shown to satisfy the condition of equal sensation jnd's. Therefore,

$$\frac{\log x}{\log k} + p\left(\frac{\log x}{\log k}\right)$$

is also a solution if $p$ is differentiable and periodic with period 1.

There is an infinity of such functions $p$, and so in infinity of different solutions to the problem for any Weber function, including Weber's law. This, of course, is quite unsatisfactory; later on we will show that one of the properties that we usually attribute to jnd's, and which as yet we have not used, enables us to insure a unique solution. However, first it will be useful to present Koenigs's results on the existence of solutions to Abel's equation.

*The existence of solutions to Abel's equation.* In psychophysical problems, there is always a threshold $R > 0$, such that $g(x)$ is not observable in the range $0 \leq x \leq R$. Thus, it is only a matter of convenience what we assume about the behavior of $g$ near 0; we shall suppose that

$$g(0) = 0 \quad \text{and} \quad 0 < g'(0) < \infty$$

where $g'(x) = \dfrac{dg}{dx}$. It is known also

from experimental work that $g$ is never 0 and that on the whole it will increase with $x$, except for limited ranges of some stimuli, where it may decrease slowly. With little or no loss of generality, we may suppose it never decreases so rapidly as to have a slope less than $-1$. In other words, we also assume:

$$g(x) > 0 \quad \text{and} \quad g'(x) > -1 \quad \text{for} \quad x > 0$$

From these assumptions, it follows that $f(x) = x + g(x)$ has these properties:

$f$ if strictly monotonic in $x$, i.e., if $x < y$, then $f(x) < f(y)$; 0 is the only fixed point of $f$ ($x$ is a fixed point if $f(x) = x$); and $1 < f'(0) < \infty$.

The strict monotonicity of $f$ implies that there exists an inverse function $f^{-1}$, i.e., a function such that

$$f^{-1}[f(x)] = x = f[f^{-1}(x)]$$

It is easy to show that:

$f^{-1}$ is strictly monotonic increasing, $x$ is a fixed point of $f^{-1}$ if and only if $x = 0$, $0 < f^{-1\prime}(0) < 1$.

Observe that if we know a solution $v$ to the equation

$$v[f^{-1}(x)] = 1 + v(x) \qquad [3]$$

then $u = -v$ is a solution to

$$u[f(x)] = 1 + u(x) \qquad [4]$$

So it will suffice to deal with $f^{-1}$. If, in addition to the three properties mentioned, $f^{-1}$ is analytic, i.e., if there exist constants $a_i$ such that

$$f^{-1}(x) = \sum_{i=0}^{\infty} a_i x^i,$$

then Koenigs has shown that a differentiable solution exists to Abel's equation. In applications, analyticity is no real restriction. For simplicity of notation, let us denote $f^{-1}$ by $h$; then Koenigs' theorem (which is not easy to prove) may be expressed as follows: Let $h^{(n)}$ denote the $n$th iterate of $h$ (i.e., $h^{(n)}(x)$ is the result of $n$ successive applications of $h$ beginning at the point $x$), and let

$$\phi(x) = \lim_{n \to \infty} \frac{h^{(n)}(x)}{[h'(0)]^n}$$

then $\phi$ exists and is differentiable, and

$$v_0(x) = \frac{\log \phi(x)}{\log h'(0)}$$

is a solution to Abel's Equation 3. Therefore, since $h'(0) = 1/f'(0)$,

$$u_0(x) = \frac{\log \phi(x)}{\log f'(0)}$$

is a solution to Equation 4 and so to our problem.

The difficult part of the proof is to show that the limit exists. Assuming that it does, it is easy to show that $u_0(x)$ is a solution. Since $h[f(x)] = x$, $u_0[f(x)]$

$$= \left\{ \log \lim_{n \to \infty} \frac{h^{(n)}[f(x)]}{[h'(0)]^n} \right\} \bigg/ \log f'(0)$$

$$= \left\{ \log \frac{1}{h'(0)} \lim_{n \to \infty} \right.$$
$$\left. \times \frac{h^{(n-1)}(x)}{[h'(0)]^{n-1}} \right\} \bigg/ \log f'(0)$$

$$= \frac{\log f'(0) + \log \lim_{n \to \infty} \dfrac{h^{(n-1)}(x)}{[h'(0)]^{n-1}}}{\log f'(0)}$$

$$= 1 + u_0(x)$$

The evaluation of the above limit for $\phi$ is rarely a simple task. Furthermore, the conditions under which it has been shown to exist and to provide a solution to Abel's equation are only sufficient conditions—there are other circumstances in which solutions exist. For example, the function $f(x) = ax^b$, $b \neq 1$, fails to satisfy $1 < f'(0) < \infty$, yet by direct verification one can show that

$$u_0(x) = \frac{\log \log [a^{1/(b-1)}x]}{\log b}$$

satisfies $u_0(ax^b) = 1 + u_0(x)$. The function $f(x) = x + ax^b$ also fails to meet the same condition, but a solution probably exists in this case too. Presumably, other functions can be found which approximate empirical data and which meet the assumed conditions, but it remains to be seen whether the limit $\phi$ can be evaluated.

The difficulty is, first, in inverting $f$, and second, in finding a simple expression for $h^{(n)}$. Since this is generally difficult, we doubt that the mathematics of this section will be useful to psychophysicists who want a nongraphic method for cumulating jnd's.

It should be pointed out again that for the empirically important Weber function $g(x) = kx + c$ the solution is known: it is

$$u_0(x) = \frac{\log (kx + c)}{\log (1 + k)}$$

*A further definition of the sensation continuum.* So far we have examined two formulations of Fechner's problem, both of which are unsatisfactory. The first, that of Fechner, contains an internal contradiction. The second, the functional equation formulation, we have shown can be solved. Unfortunately, we have also shown that it has infinitely many families of different solutions, which is intolerable. In this section we shall propose an addition to the second formulation which amounts to a method of summating jnd's. We shall show that if we demand a particular form of invariance of distances measured in jnd units, then there is a unique (except for zero and unit) sensation scale for each of a wide variety of Weber functions, and for Weber's law this sensation scale is Fechner's law.

The common psychological custom for measuring distances in jnd's between two points is to use the size of the jnd at the lower point as the unit of measurement. Although it is rarely if ever explicitly stated, it is certainly implicitly assumed that if the distances $ab$ and $cd$ are both $\alpha$ stimulus jnd's in length, then they have an equal number, say $K(\alpha)$, of sensation jnd's. As a formal mathematical condition, this states that

$$u[x + \alpha g(x)] - u(x) = K(\alpha)$$

where $K$ is some fixed, but unknown, function of $\alpha$. It can be shown, first, that if $u$ is a solution to this problem, then it must be the integral $\int \dfrac{dx}{g(x)}$ given by Fechner, but, second, that there are no solutions except when $g(x) = cx$ (Weber's law). We will not present a proof of this result since it is a blind alley, but we believe that it suggests that this customary measurement of distances should be abandoned.

We must now consider how such distances really should be measured. If $x$ and $y$ are more than one jnd apart, we may expect the size of the jnd to change as we go from $x$ to $y$. That fact should be taken into account in using jnd's as units of measurement; failure to take it into account is what makes Fechner's auxiliary principle and the standard measuring procedure unacceptable. We shall proceed to formulate this more sensible method of using jnd's as measuring units.

Let $f(x) = x + g(x)$; then the point $f(x)$ is one $x$-jnd larger than $x$. The point $f[f(x)] = f^{(2)}(x)$ is one $f(x)$-jnd larger than $f(x)$. In general $f^{(n)}(x)$ is one $f^{(n-1)}(x)$-jnd larger than the point $f^{(n-1)}(x)$. Clearly, for $y > x$, we can find some integer $n$ such that

$$f^{(n)}(x) \leq y < f^{(n+1)}(x)$$

and it is reasonable to say that $y$ is between $n$ and $n + 1$ jnd's larger than $x$. For the moment, let us suppose that $y$ was chosen so that $y = f^{(n)}(x)$, then we can say $y$ is exactly $n$ jnd's larger than $x$. It seems plausible to require that the same be true of the sensory continuum, i.e.,

$$u[f^{(n)}(x)] - u(x) = n$$

In words, we are saying that if point $y$ is 20 stimulus jnd's higher than

point $x$ on the stimulus continuum, then it must also be 20 sensation jnd's higher than point $x$ on the sensation continuum. If the above condition is met for $n = 1$ (in other words, if all sensation jnd's for a given sensory continuum are equal), then it must also be met for all larger values of $n$, since

$$u[f^{(n)}(x)] - u(x)$$
$$= u\{f[f^{(n-1)}(x)]\} - u(x)$$
$$= 1 + u[f^{(n-1)}(x)] - u(x)$$
$$\cdots$$
$$= n$$

But this takes care of relatively few points, and does not allow us to say exactly how many jnd's $y$ is from $x$ unless the difference is a whole number of jnd's. We must find a definition which tells us how to subdivide a jnd into fractional parts. How to do this is not obvious, since the definition of distances given above involves iterates of $f$, and these are apparently defined only for integers. Fortunately, it is possible to generalize the notion of an iterate to arbitrary, rather than integral, indices. This problem is closely related to that of Abel's functional equation which Koenigs examined; we shall be able to use his results.

First, we can set up some properties that a generalized iterate $f^{(t)}(x)$, where $t$ is any non-negative number, should meet. In essence, they amount to stipulating that $f^{(t)}(x)$ should coincide with the usual definition when $t$ is an integer and that the same law of composition should hold. Formally, it is sufficient to require that

$$f^{(0)}(x) = x, \quad f^{(1)}(x) = f(x)$$

and for every $s$ and $t \geq 0$,

$$f^{(s+t)}(x) = f^{(s)}[f^{(t)}(x)]$$

For integers, the generalized iterate coincides with the usual notion, as you can see, by repeatedly applying the last condition to the second one.

We have already presented a result of Koenigs which showed that if $f$ is strictly monotonic and analytic, $1 < f'(0) < \infty$, and 0 is the only fixed point of $f$, then there exists a function $\phi$ defined in terms of the iterates of $f^{-1}$ such that

$$u_0(x) = \frac{\log \phi(x)}{\log f'(0)}$$

is a basic solution to Abel's equation. This means that $\phi$ is itself a solution to what is called Schroder's equation

$$v[f(x)] = f'(0)v(x)$$

which is obtained from Abel's by taking exponentials on both sides. Using this fact and following Koenigs, it is easy to show that $\phi^{-1}$ exists and that the function

$$f^{(t)}(x) = \phi^{-1}\{[f'(0)]^t\phi(x)\}$$

satisfies the three conditions of a generalized iterate. We show the latter. First,

$$f^{(0)}(x) = \phi^{-1}[\phi(x)] = x$$

Second,

$$f^{(1)}(x) = \phi^{-1}[f'(0)\phi(x)]$$

And so, using the fact that $\phi$ satisfies Schroder's equation,

$$\phi[f^{(1)}(x)] = f'(0)\phi(x) = \phi[f(x)]$$

Hence,

$$f^{(1)}(x) = f(x)$$

Finally,

$$f^{(s)}[f^{(t)}(x)]$$
$$= \phi^{-1}[[f'(0)]^s\phi(\phi^{-1}\{[f'(0)]^t\phi(x)\})]$$
$$= \phi^{-1}\{[f'(0)]^s[f'(0)]^t\phi(x)\}$$
$$= \phi^{-1}\{[f'(0)]^{s+t}\phi(x)\}$$
$$= f^{(s+t)}(x)$$

So, with this definition of the generalized iterate, we can generalize the above definition of distances in jnd's to prescribe how to deal with fractional jnd's.

We reformulate our major mathematical problem:

Given a Weber function $g$ which is analytic, $g'(x) > -1$ for all $x > 0$, $g'(0) > 0$, and $g(0) = 0$, to find those functions $u(x)$ such that $u[f^{(t)}(x)] - u(x) = t$, for all $x > 0$, and all $t > 0$, where $f^{(t)}$ is the generalized iterate of $f(x) = x + g(x)$.

Note that by setting $t = 1$, this condition implies the equality of sensation jnd's.

First, we show that $u_0(x) = \dfrac{\log \phi(x)}{\log f'(0)}$

solves the reformulated problem:

$$u_0[f^{(t)}(x)] - u_0(x)$$

$$= \frac{\log \phi[f^{(t)}(x)]}{\log f'(0)} - \frac{\log \phi(x)}{\log f'(0)}$$

$$= \frac{\log \{[f'(0)]^t \phi(x)\} - \log \phi(x)}{\log f'(0)}$$

$$= \frac{t \log f'(0)}{\log f'(0)}$$

$$= t$$

Second, from the results about Abel's equation, we know that if there are any other solutions to this problem, they must be of the form $u_p = u_0 + p(u_0)$, where $p$ is periodic with period 1. For $u_p$ actually to solve the reformulated problem, it is necessary that for every $t \geq 0$,

$$t = u_p[f^{(t)}(x)] - u_p(x)$$

$$= u_0[f^{(t)}(x)] + p\{u_0[f^{(t)}(x)]\}$$

$$\qquad - u_0(x) - p[u_0(x)]$$

$$= t + p[t + u_0(x)] - p[u_0(x)]$$

Thus, for every $t \geq 0$,

$$p[t + u_0(x)] = p[u_0(x)]$$

That is, $p$ must be periodic with every period $t$, and so $p$ is a constant. Thus, up to an additive constant, $u_0$ is the unique function which solves our reformulated problem.

In nonmathematical language, introducing the method of measuring fractional jnd's has enabled us to eliminate all solutions to Abel's equation save $u_0$, thus cutting down the number of acceptable solutions from infinity to one.

We conclude, therefore, that the condition stated in the reformulated problem constitutes an acceptable definition of a psychophysical sensation continuum, in the sense that it yields a unique Fechner function for any reasonable Weber function. We also find that for Weber's law this condition yields Fechner's law. The solution of our reformulated problem may cause unhappiness because it is not the same as the integral "solution" proposed by Fechner, except in the special case of the linear generalization of Weber's law. However, we have already shown that the integral "solution" contradicts the definition of equal sensation jnd's.

It is sad that the integral is not the right solution, for its evaluation is often easy, and we fear that no working psychophysicist will find in our mathematics a tool for determing a summated jnd scale any better or more efficient than the simple graphic procedure of adding jnd's up one at a time.

*The statistical nature of jnd's.* So far we have sounded as though we were treating jnd's as fixed quantities, although every psychophysicist knows that jnd's are statistical fictions, defined by an arbitrarily chosen cutoff on a cumulative frequency curve.

However, we now show that our method of reducing the infinity of solutions to Abel's equation to one is equivalent to treating jnd's as just such statistical fictions.

We start with the old, famous psychological rule of thumb: equally often noticed differences are equal, unless always or never noticed. We define $P(y,x)$ as the probability that $y$ is discriminated as larger than $x$. Now, this rule of thumb simply means that on the sensation continuum the function $P(y,x)$ is transformed in such a way that it no longer depends on $x$ and $y$ separately, but only on the difference of their transformed values. Put another way, the subjective continuum $u$ is a strictly monotonic transformation of the stimulus continuum such that the probability that a change of $\delta$ units on the sensation scale will be detected depends only upon $\delta$, and not on the place at which $\delta$ begins or ends.

Formally, if we are at a point $x$ of the stimulus continuum, and therefore at $u(x)$ on the sensation scale, and if a stimulus $y$ is presented such that $u(y) = u(x) + \delta$, then the chance that $y$ will be detected depends upon $\delta$, but not on $x$. If we note that

$$y = u^{-1}[u(x) + \delta]$$

then the condition is that

$$P\{u^{-1}[u(x) + \delta], x\} = P(\delta)$$

Our problem is to decide under what conditions this problem has a solution and what that solution is. To this end, we make the assumption that for each $x$, $P(y,x)$ is a strictly monotonic increasing function of $y$.

We show the following: If the above problem has a solution, then there exists a function $f(x)$ such that $P[f^{(t)}(x), x]$ is independent of $x$, where $f^{(t)}(x)$ is the $t^{th}$ iterate of $f(x)$ previously defined. The function

$f(x) - x$ is a Weber function naturally defined in terms of $P$. If there is a solution, it is unique and it is the solution $u_0$ to Abel's equation $u[f(x)] - u(x) = 1$. In other words, if there is any solution to the problem of equally often noticed differences being equal, then it is unique and it is the solution to our proposed reformulation of Fechner's problem.

The proof is comparatively simple and runs as follows. Suppose there exists a solution $u$ to the condition that $P\{u^{-1}[u(x) + \delta], x\}$ is independent of $x$ for all $\delta \geq 0$. Since $P$ is strictly montonic in $y$ for all $x$, there is a unique solution to $P(y,x) = k$ for each $k$, $0 < k < 1$; call it $y = f_k(x)$. For any $\delta$, let $k = P(\delta)$, and so by our assumption $u$ must satisfy

$$u^{-1}[u(x) + \delta] = f_\delta(x)$$

where we have written $f_\delta$ for $f_{P(\delta)}$. Applying $u$ to this, we have

$$u[f_\delta(x)] - u(x) = \delta$$

Let $f = f_1$. We observe that if $\delta = 0$, then $f_0(x) = x$. Suppose we choose any $\delta, \epsilon \geq 0$ and let $y = f_\epsilon(x)$; then

$$\begin{aligned}
\delta &= u[f_\delta(y)] - u(y) \\
&= u\{f_\delta[f_\epsilon(x)]\} - u[f_\epsilon(x)] \\
&= u\{f_\delta[f_\epsilon(x)]\} - u(x) - \epsilon
\end{aligned}$$

Thus,

$$u\{f_\delta[f_\epsilon(x)]\} - u(x) = \delta + \epsilon$$

But, from above,

$$u[f_{\delta+\epsilon}(x)] - u(x) = \delta + \epsilon$$

so

$$u[f_{\delta+\epsilon}(x)] = u\{f_\delta[f_\epsilon(x)]\}$$

whence

$$f_{\delta+\epsilon}(x) = f_\delta[f_\epsilon(x)]$$

Thus, we have shown that $f_\delta$ must satisfy the three conditions of a gen-

eralized iterate of $f$, i.e., $f_\delta = f^{(\delta)}$ for all $\delta$, so a necessary condition for a solution is that

$$P[f^{(\delta)}(x), x]$$

shall be independent of $x$. From the fact that $u[f_\delta(x)] - u(x) = \delta = u[f^{(\delta)}(x)] - u(x)$, it follows that the solution is uniqe and that it is the same as that given for our reformulation of Fechner's problem.

It probably is not obvious, but the point of this section extends beyond sensory psychophysics into the scaling procedures based on Thurstone's law of comparative judgment. Case V of that law is based on the assumption that equally often noticed differences are equal unless always or never noticed. This fact has two interesting implications. The first and more obvious one is that these two apparently different branches of psychological measurements are actually doing the same thing (namely, using a measure of confusion as a unit of measurement by assuming that confusion is equal at all places on the subjective scale). The second, less obvious implication is that perhaps sensory psychophysics can profit by considering, as Thurstone and his followers have, scaling methods with less rigid assumptions which nevertheless are based on confusability data. One of us (Luce) will pursue this possibility further in a forthcoming book (7).

*Graphic methods for cumulating jnd's.* Psychophysical data do not come in mathematical form. In order to apply our method for cumulating jnd's (or Fechner's, for that matter), it is necessary either to put the Weber function into equation form, or else to develop a graphic equivalent of the appropriate mathematical operations. The graphical equivalent of Fechner's technique is well known, although rarely used (see, e.g., **15**, pp. 94 and
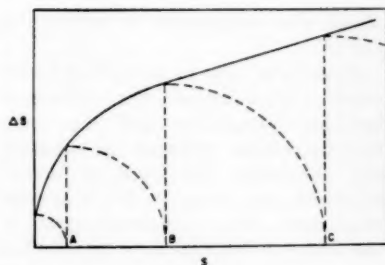


FIG. 1. How to cumulate jnd's. The size of the jnd at the origin is marked off on the $x$ axis to find point $A$, the size of the jnd at $A$ is marked off to find point $B$, and so on. The stimulus values $A$, $B$, $C$, . . . correspond to the points 1, 2, 3, . . . on the cumulated jnd scale.

147–148). It is, of course, wrong, since Fechner's technique is wrong. If our technique is to be of greatest applicability, we should provide a graphic equivalent also. Unfortunately, it seems difficult to find a truly convenient one. The only method we know of is to go back to the basic idea of adding up jnd's—the idea that one jnd plus one jnd is two jnd's. The method of applying this basic idea is given in Figure 1, and was discussed earlier in the paper. Its error characteristics are abour the same as those of the graphic techniques of integration which have been used in the past. Unfortunately, the method is tedious; if there are 170 jnd's between absolute threshold and the upper limit of discrimination, then 170 separate operations are required to determine the cumulated-jnd scale. The errors in these successive operations do not multiply, however.

*Practial effects of the new procedure.* No doubt it is important to understand Fechner's logical error and to know how to avoid it, but the burning question for working psychophysicists is: What, if anything, does this do to the currently accepted conclusions

about the uselessness of adding up jnd's?

First, it is easy to show that under some circumstances the difference between integration and the functional-equation solution is substantial. Consider the class of Weber functions $g(x) = ax^{1+e}$: if $e$ is greater than zero, the asymptotic error of the integral solution as $x$ approaches infinity is infinite; while if $e$ is less than zero, the asymptotic error is zero. Of course, if $e$ equals zero (Weber's law), the two procedures give identical results. The order of magnitude of the error for small numbers of jnd's depends on the constants in the equation; it can be of significant size even if $e$ is less than zero. One way of looking at it is that the integral solution is the approximation given by the first two terms of a Taylor series expansion of the functional equation; all square and higher power terms of the expansion are omitted:

$$u[x + g(x)] - u(x) = 1$$
$$= u(x) + u'(x)g(x)$$
$$\quad + u''(x)\frac{g(x)^2}{2!} + \ldots - u(x)$$
$$= u'(x)g(x)$$
$$\quad + \left\{ u''(x)\frac{g(x)^2}{2!} + \ldots \right\}$$

A number of experimental determinations of jnd's, particularly for intensive continua, produce a curve of $g(x)/x$ that first falls and then is flat —a function often well approximated by $g(x) = kx + c$. However, for some continua the picture is less simple. There are some (pitch, for example) where the curve appears to rise again at the high end. The falling section of these curves corresponds to the case $e < 0$; the flat section corresponds to the case $e = 0$; the rising section corresponds to the case $e > 0$.

However, the $x$-axis of such graphs is usually plotted logarithmically. This means that the rising section may cover most of the range within which the stimulus can be varied—a fact which the logarithmic $x$-axis tends to conceal. So it is quite possible that the error in using the integration technique is substantial for many sense modalities and for large ranges within each.

But the possibility of error is irrelevant unless someone has actually made the error. Has anyone? Extensive examination of the literature suggests that the answer is that not very many such errors have occurred. Some authors are quite unclear about how they added up jnd's, but many of them have preferred the step-by-step method which corresponds to the functional-equation solution because it was very simple to do. How simple it is, of course, depends on the number of jnd's to be added; we doubt very much if the jnd's for pitch will ever be added this way, since there are several thousand of them. We have found only one clear instance (15) in which the graphic equivalent of integration has been used (to cumulate pitch jnd's, as it happens), though it has been vigorously recommended. The general avoidance of the graphic equivalent of integration may be caused by shrewd intuition that something is wrong with Fechner's mathematical auxiliary principle. Or it may simply be a rare instance in which the fear of mathematical complexity has benefited science.

*Do cumulated jnd's agree with other scales?* The results of cumulating jnd's have often been compared with the results of other psychophysical procedures (4). The most common finding has been that the cumulated jnd scales do not agree with scales de-

termined by fractionation or direct magnitude estimation, at least for such continua as loudness. A review of this literature might seem appropriate here, but it is quite unnecessary, since the relation between scales based on confusion data (like cumulated jnd scales) and those based on fractionation or magnitude estimation has been extensively and excellently discussed in recent studies by Stevens (**14**), Stevens and Galanter (**16**), and Piéron (**8, 9, 10**).

The controversy over the relation between cumulated jnd scales and scales determined by other methods is embedded in a larger, sometimes acrimonious controversy about the relationships among various methods of sensory scaling. To some extent we shall have to enter the fray.

The first and most important question is this: Do the different scaling procedures, if properly used, lead to different scales? Unless we reject a great many experiments as improperly performed, we must answer "Yes." But the issue is not as simple or unambiguous as that answer. For example, Garner (**3**) has developed a loudness scale based on both fractionation and multisection judgments that fits a large number of experimental results in auditory psychophysics better than does the old sone scale (his paper was written prior to the development of the new sone scale [**13**]). Figure 2 shows the relationship between that scale and a cumulated jnd scale for loudness prepared by us from Riesz's data (**11, 12**). The two scales seem to be roughly linearly related—but does it mean anything for the controversy? Riesz's procedure has often been criticized, and his data are almost 30 years old. The form of Garner's scale (which is all that matters for this argument) is based primarily on his multisection
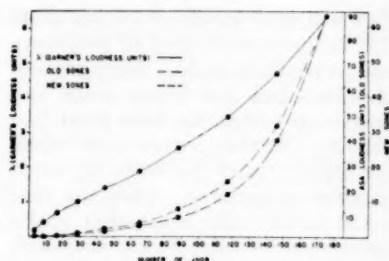


Fig. 2. The relation between Garner's loudness scale and Riesz's cumulated jnd scale. The old sone scale (ASA loudness scale) and Stevens's recent revision of it are included for comparison.

rather than his fractionation data. Scales based upon multisection data usually agree with those constructed by confusability methods; the explanation proposed by critics of these methods is that the adjustment of five or six stimuli in a multisection experiment may produce confusion among the tones being adjusted. If this argument is correct, and if the form of Garner's scale is based upon multisection data, it is not surprising that the two agree. Our reason for so extensive a discussion of Garner's scale is that loudness is the central battleground of this controversy. If the verdict of psychophysical history is that confusability and multisection scales give results different from fractionation results for loudness, then psychologists will amost certainly assume that the two procedures yield different results in other intensive (or, as Stevens calls them, Class I or prothetic) continua. Unfortunately, even in psychophysics, not enough universally accepted data are available to settle the argument.

If confusability scales and scales based upon fractionation or direct magnitude estimation agree, no problem arises. If not (and we suspect

they will not), psychophysicists must still evaluate each kind of procedure and its resulting scale. Some psychophysicists feel that fractionation and magnitude estimation have great face validity, and that confusability scales are distortions of the scales obtained by these procedures. They say that fractionation and estimation scales correspond to what $S$s say they feel, they are obtained by straightforward procedures rather than indirect ones, and, after all, what logic is there in basing a measure of magnitude on variance or "noise."

Other psychophysicists feel that confusability scaling is the better method. They say that fractionation and estimation data are unreliable, variable, and, as a rule, at least fractionation data cannot be turned into scales unless obtained from a "good," which means extensively trained, subject. The estimation techniques have not been used enough times in enough places to indicate clearly what effect, if any, training may have on the results. Confusability scales can be obtained from untrained $S$s who have no idea what form of scale is wanted from them; they can even be obtained from animals.

Each group asserts that its preferred scales are more nearly consistent with the bulk of psychophysical data than the other kind of scales; each group can produce impressive arguments to buttress its claim.

Still another position is possible: perhaps two different kinds of sensory processes are being tapped by these two different kinds of procedures. If so, both kinds of scales are useful, but for different purposes. This could well be the eventual endpoint of the argument.

Yet another source of confusion in the argument is the treatment of individual differences. The custom has

been to take means or medians, and recently a number of psychophysicists have raised vigorous questions about the appropriateness of doing so. W. J. McGill[5] is currently attempting to find a better way of respecting individual differences while still obtaining a "universal" scale. It will be interesting to see what light serious attempts to do justice to individual differences sheds on the differences between the two classes of scales.

The status of cumulated jnd's has been controversial for more than a hundred years, and this paper is not intended as an attempt to settle the controversy. Our main point is that Fechner's problem has been improperly formulated and that the integral usually offered as a solution is not in fact a solution when the Weber function differs from $g(x) = kx + c$. We have also developed what appears to be the correct solution, only to find that in computational work it has usually been used in spite of its disagreement with the integral solution. This means that our clarification of the logical issues underlying Fechner's formulation does little to change the status of the present, primarily empirical, controversy about scaling methods. However, one of us (Luce [7]) has recently developed a way of dealing with confusability data based on a simple axiom which, if it works out successfully, may resolve the difficulty by changing our ideas about the meaning of confusability scales; this development will be described in another publication.

## SUMMARY

Fechner's method for adding up just noticeable differences (jnd's) to obtain sensory scales is based on a

mathematical error: he used a differential equation approximation to a functional equation instead of the functional equation itself. The functional equation can, however, be solved directly. The solution coincides with the differential equation solution only in the special case in which the linear generalization of Weber's law holds exactly. The mathematical properties of the formal solution are such that it probably will not be very useful for practical computation, but the extremely simple graphical procedure of adding up jnd's one at a time is the graphical equivalent of the mathematically correct solution. The amount of difference between the two procedures can be calculated for some special cases; its size depends on the form of the function relating size of jnd's to stimulus magnitude.

This error does not seem to have any significant impact upon the controversy over the relation between cumulated jnd scales and scales based on fractionation and direct estimation data because most psychophysicists have, in fact, ignored the recommended (incorrect) procedure and have stubbornly summated jnd's in the obvious and correct way.

## REFERENCES

1. BORING, E. G. *A history of experimental psychology.* (2nd ed.) New York: Appleton-Century-Crofts, 1950.
2. FECHNER, G. T. *Elemente der Psychophysik.* (Reprint) Leipzig: Breitkopf und Härtel, 1889.
3. GARNER, W. R. A technique and a scale for loudness measurement. *J. acoust. Soc. Amer.*, 1952, **24**, 153–157.
4. HALSEY, R. M., & CHAPANIS, A. Luminance of equally bright colors. *J. optical Soc. Amer.*, 1955, **45**, 1–6.
5. KOENIGS, M. G. Recherches sur les intégrales de certaines équations fonctionnelles. *Ann. Scientifiques de l'Ecole Normale Supérieure* (3), 1884, **1**, Supplement, S1–S41.
6. KOENIGS, M. G. Nouvelles recherches sur les équations fonctionnelles. *Ann. Scientifiques de l'Ecole Normale Supérieure* (3), 1885, **2**, 385–404.
7. LUCE, R. D. *Individual choice behavior: A theoretical analysis.* New York: Wiley, in press.
8. PIÉRON, H. L'évaluation des sensations. *Bull. Psychol., Univer. Paris,* 1950, **4**, 3–38.
9. PIÉRON, H. *Les problèmes fondamentaux de la psychophysique dans la science actuelle.* Paris: Hermann, 1951.
10. PIÉRON, H. *The sensations, their functions, processes and mechanisms.* New Haven: Yale Univer. Press, 1952.
11. RIESZ, R. R. The differential intensity sensitivity of the ear for pure tones. *Physical Rev.*, 1928, **31**, 867–875.
12. RIESZ, R. R. The relationship between loudness and the minimum perceptible increment of intensity. *J. accoust. Soc. Amer.*, 1933, **4**, 211–216.
13. STEVENS, S. S. The measurement of loudness. *J. acoust. Soc. Amer.*, 1955, **27**, 815–829.
14. STEVENS, S. S. On the psychophysical law. *Psychol. Rev.*, 1957, **64**, 153–181.
15. STEVENS, S. S., & DAVIS, H. *Hearing: Its psychology and physiology.* New York: Wiley, 1938.
16. STEVENS, S. S., & GALANTER, E. H. Ratio scales and category scales for a dozen perceptual continua. *J. exp. Psychol.*, 1957, **54**, 377–411.

# STRENGTH OF CARDIAC CONDITIONED RESPONSES WITH VARYING UNCONDITIONED STIMULUS DURATIONS [1]

## NORMA WEGNER [2] AND DAVID ZEAMAN

*University of Connecticut*

The effects of long and short shocks on the conditioning of fear have been frequently considered important tests of theory of how fear is learned (1, 2, 3, 4, 9, 10). N. E. Miller, for example, has stated that according to a strict drive-reduction theory of reinforcement ". . . other things equal, a signal followed by a brief noxious stimulus should acquire the capacity to elicit stronger fear than one followed by a prolonged noxious stimulus" (2, p. 375).

Despite its theoretical significance, there have been few direct experimental attacks on this problem. We have found only two published conditioning studies with duration of noxious US as a major variable, one by Bitterman, Reed and Krauskopf (1) and the other by Mowrer and Solomon (4). Each of these studies examined only two values of shock duration over a relatively narrow range.

More extensive data on this problem can be extracted from a series of cardiac conditioning experiments we have done exploring the effects of duration of shock on the form of the conditioned heart response (8, 9, 10). These studies provide us with an assemblage of data, as yet unreported, on the relation of shock duration to strength rather than form of CR.

If heart rate change during anticipation of a shock is accepted as an index of fear, our classical conditioning stud-

ies using shocks of 0.1, 2.0, 6.0, and 15 sec. allow us to make a direct and thorough empirical check on Miller's statement of the drive reduction view. We shall first summarize our own experiments.

## METHOD

### General

The apparatus, procedure, and *S*s have all been described in detail in previous papers (8, 9, 10). They are summarized as follows. Forty-three male and twenty female college student *S*s were assigned unsystematically to four groups, each run under a classical trace-conditioning procedure with a different shock US duration (0.1, 2.0, 6.0, and 15 sec.). All groups received initially at least ten well distributed preconditioning trials of the CS alone —a 1-sec. tone (60 db, 512 c.p.s.). Following these were at least ten spaced conditioning trials consisting of the tone CS followed in 6 sec. by the shock US (13 V.A.C.). This mild but reportedly unpleasant shock was applied across the first two fingers of the left hand. Heart rate was measured with an electrocardiograph before, during, and after each trial.

### Measures of CR and UR

Since individual differences in the form of conditioned cardiac responses occur, a measure of heart rate disturbance was chosen which would be comparable for the different types of response. This was a variability measure, the statistical *range* of heart rates during the conditioning (posttone) period. It was simply derived as follows. For a single *S* on every conditioning trial each heart beat interval between the tone and shock was converted to a rate measure. There were usually about eight or nine of these. The difference between the largest and smallest of them defined the *range* for that trial, a measure which we previously have designated as *Maximum CR*. A mean is taken over the first ten conditioning trials for this measure to represent a single *S*.

The same steps as above are taken to compute the *Maximum UR* except that ten beats

after the shock onset are used instead of the eight or nine before.

## RESULTS

Figure 1 presents the Maximum CR and Maximum UR for four groups having different shock durations. Each point is the mean of 15 Ss' data (3 Ss, selected without bias, were omitted to equalize the Ns in the groups for statistical convenience). Maximum CR does not seem to vary systematically as a function of shock US duration. What fluctuations do occur are easily attributable to the corresponding variation in Maximum UR. This fact is established by Table 1, which gives the results of an analysis of covariance of the Maximum CR function, using Maximum UR as the relevant covariate. When the influence of the unconditioned response is extracted from the conditioned response there is left an insignificant amount of CR variation among the shock conditions.

The absence of significant differences among the group means is not to be interpreted as due to absence of conditioning. Evidence has already been presented (8, 9, 10) to show that a significant amount of cardiac conditioning
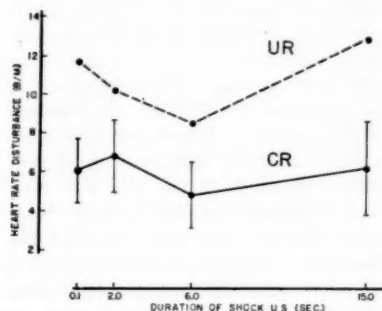


FIG. 1. The maximum amplitudes (range) of CR and UR are plotted as a function of shock US duration. One standard deviation is indicated by a vertical line on either side of each of the mean values comprising the points of the CR function.

### TABLE 1

SUMMARY OF ANALYSIS OF COVARIANCE OF MAXIMUM CR AMPLITUDE WITH MAXIMUM UR AMPLITUDE

| Source | $df$ | $M^2$ | $F$ |
|---|---|---|---|
| Between durations | 3 | 766 | .68 |
| Within | 52 | 1131 | |

For 3 and 52 $df$, $F$ is 2.79 for $p = 5\%$.
The variate-covariate correlation is $+.57$.

took place in all four groups, using other (correlated) measures of response. To demonstrate further that the CRs and URs in Fig. 1 represent reliable disturbances in rate, we have compared them with similar measures of pretone variability, i.e., the range of ten pretone rate measures. Table 2 presents the results of a series of $t$ tests of the significance of the differences between mean pretone range and mean posttone range (for CR), and between mean pretone range and postshock (for UR). Each $t$ is based on 15 differences, each difference itself a mean (over 10 trials) for each of 15 Ss. Pretone means were small, less than 1.5 B/M for all groups, in comparison to the posttone and postshock means shown in Fig. 1.

## DISCUSSION

We wish to discuss our results briefly in relation (a) to previous findings and (b) to Miller's theoretical statement.

### TABLE 2

PROBABILITY OF CHANCE OCCURRENCE OF MEAN DIFFERENCES BETWEEN PRETONE CARDIAC DISTURBANCE AND BOTH CR AND UR FOR THE FOUR DURATIONS OF SHOCK US

| | Shock Duration (Sec.) | | | |
|---|---|---|---|---|
| | 0.1 | 2.0 | 6.0 | 15 |
| CR | .01 | .01 | .02 | .01 |
| UR | .001 | .001 | .01 | .001 |

The flat gradient in Fig. 1 relating strength of CR to US duration is not inconsistent with the findings of Bitterman, Reed, and Krauskopf (1), who found no significant difference in the strength of conditioned GSR with shock duration of .5 and 3 sec. This lack of difference, however, may have appeared because they used the same $Ss$ for both durations. Their experiment employed different signals for short and long shocks; and, if the $Ss$ discriminated the signals verbally, they were assumed not to have generalized autonomically (i.e., GSR). The weakness of this assumption is shown by the fact, reported by Notterman, Schoenfeld and Bersh (5), that $Ss$ who state they expect no shocks during experimental extinction continue to show conditioned cardiac disturbance.

Mowrer and Solomon (4) paired a light signal with shocks of 3 sec. for one group of rats and with shocks of 10 sec. for another group. The strength of fear conditioned to the light was measured by the capacity of the light to act as effective punishment of a lever-pressing response. For these two durations of shock, an abrupt onset and termination of shock was employed. No significant difference between the two groups was found with respect to their inferred fear. Two further groups were run under a similar procedure, except that shock was terminated gradually rather than abruptly. For one group the shock was 4 sec. long and for the other, 7 sec. No reliable difference in fear was reported between these groups either. Despite the wide differences in $Ss$, procedures, and methods of inferring fear, the results of the Mowrer and Solomon study are in agreement with our own findings. Such agreement gives us confidence that there is no relationship between indices of fear and noxious stimulus durations.

The discrepancy between the empirical findings and Miller's theoretical formulation of the drive-reduction view deserves comment. It will be noted that Miller prefaced his statement that short noxious stimuli would produce more fear than long ones with the phrase, "other things equal." There are at least two important "other things" here; the first is strength of US, or more precisely, *subjective* strength. This amounts to a statement of the need to control the strength of the shock, or other noxious stimulus, as it feels to the subject. If, through temporal summation, the longer shocks are felt as stronger, then stronger fear might result from them despite the correctness of Miller's statement, because more intense unconditioned stimuli are known to create stronger CRs (6, 7). The second important factor in need of control is adaptation to shock. If there is a period of rapid adaptation during the early part of a prolonged shock, it might produce the major part of drive reduction necessary to condition fear. An effect of this kind shortly after shock onset could make the length of the physical shock quite irrelevant to the timing of the drive reduction.

With respect to the first point, temporal summation, we have one way of assessing the apparent or subjective intensity of shock, that is, by the maximum or total magnitude of the unconditioned response. Our results show that there were no systematic differences in the magnitude of UR for the different shock durations. Hence, we have some evidence, at least, that the over-all subjective intensity of the US was relatively constant for the various durations. We conclude from this that the discrepancy between Miller's theoretical statement and our data can not easily be attributed to lack of control of shock intensity.

With respect to the second factor, shock adaptation, we appeal this time to the form of the unconditioned response as a relevant datum. It is true that

adaptation to the shock occurs if momentary magnitude of the UR reveals subjective shock intensity. We have shown (8, 9, 10) that *regardless of shock duration* in the 0.1 − 15 sec. range the heart-rate response to shock takes the form of a rapid acceleration to a maximum rate within 4 beats of shock onset, followed by a slow irregular return (adaptation) to normal requiring more than 20 beats. If the drive property of shock parallels the course of UR magnitude in time, we would be led to say that drive reduction is remarkably unrelated to shock duration in the range investigated. In terms of Miller's application of Hull's theory, this would mean that no differences in conditioning of fear should occur because subjective shock intensity reduction (the presumed reinforcing agent) was equally delayed for all groups. If this line of reasoning is correct, our experiment unfortunately provides no real test of a theory that predicts the consequences of varying delays of reinforcement.

This difficulty, moreover, may not be restricted to the present experiment. As long as there is no demonstrable relationship between objective and subjective properties of shock, the possibility of a simple experimental test of Miller's assertion is precluded. Use of a stronger shock (if Ss would endure long durations) might conceivably eliminate the troublesome adaptation effect, but only at the cost of a corresponding increase in the likelihood of destroying our present control of temporal summation.

The present data may be accounted for by two-factor or contiguity theories which stress the importance of the onset rather than the offset of shock (3, 4), but the data are not critical tests of this type of theory.

## SUMMARY

No relationship has been found in human Ss between conditioned heart rate response magnitude and a wide range of shock US durations.

Under the assumption that cardiac disturbance is an index of fear, this fact is related to a deduction from a drive-reduction theory of fear, according to which a signal followed by a brief noxious stimulus should require the capacity to elicit stronger fear than one followed by a prolonged noxious stimulus. It is concluded that this proposition is either incorrect or not testable under the conditions provided.

## REFERENCES

1. BITTERMAN, M. E., REED, P., & KRAUS-KOPF, J. The effect of the duration of the unconditioned stimulus upon conditioning and extinction. *Amer. J. Psychol.*, 1952, **65**, 256–262.
2. MILLER, N. E. Comments on multiple-process conceptions of learning. *Psychol. Rev.*, 1951, **58**, 375–381.
3. MOWRER, O. H. Two-factor learning theory: Summary and comment. *Psychol. Rev.*, 1951, **58**, 350–354.
4. MOWRER, O. H., & SOLOMON, L. N. Contiguity vs. drive-reduction in conditioned fear: The proximity and abruptness of drive-reduction. *Amer. J. Psychol.*, 1954, **67**, 15–25.
5. NOTTERMAN, J. M., SCHOENFELD, W. N., & BERSH, P. J. Conditioned heart rate response in human beings during experimental anxiety. *J. comp. physiol. Psychol.*, 1952, **45**, 1–8.
6. PASSEY, G. E. The influence of intensity of unconditioned stimulus upon acquisition of a conditioned response. *J. exp. Psychol.*, 1948, **38**, 420–428.
7. SPENCE, K. W. Learning and performance in eyelid conditioning as a function of intensity of the UCS. *J. exp. Psychol.*, 1953, **45**, 57–63.
8. ZEAMAN, D., DEANE, G., & WEGNER, NORMA. Amplitude and latency characteristics of the conditioned heart response. *J. Psychol.*, 1954, **38**, 235–250.
9. ZEAMAN, D., & WEGNER, NORMA. The role of drive reduction in the classical conditioning of an autonomically mediated response. *J. exp. Psychol.*, 1954, **48**, 349–354.
10. ZEAMAN, D., & WEGNER, NORMA. A further test of the role of drive reduction in human cardiac conditioning. *J. Psychol.*, 1957, **43**, 125–133.

# STIMULUS AND RESPONSE GENERALIZATION: DEDUCTION OF THE GENERALIZATION GRADIENT FROM A TRACE MODEL [1]

ROGER N. SHEPARD

*Harvard University*[2]

It is now generally acknowledged that (*a*) a response conditioned to one stimulus tends also to occur to other stimuli, and (*b*) the magnitude of this response-tendency (for any particular one of those stimuli) is governed by the dissimilarity between that stimulus and the stimulus to which the response was originally conditioned. Indeed, this principle of stimulus generalization is of such fundamental importance that any quantitative theory of behavior that fails to deal with it explicitly can only be regarded as incomplete. It is not surprising, therefore, that a number of investigations have been specifically concerned with the form of the function relating generalized response-tendency to interstimulus dissimilarity. What may seem surprising, though, is that the conclusions of these various studies, far from converging on a unique function, have diverged to the several different functions illustrated in Fig. 1. Beyond those who doubt the

existence of a quantitatively invariant "gradient of generalization" in the first place (**5, 16, 21**), there are those, like Schlosberg and Solomon, who consider this gradient to be linear, as in A (**8, 14, 22, 23**); those, like Spence, who consider it to be convex upward, as in B (**27, 28**); and those, like Hull, who consider it to be concave upward, as in C (**1, 11, 12, 19, 20**). In addition, if the "discriminal dispersion" of the psychophysicists can be interpreted as a gradient of generalization, there is support for a bell-shaped function which is first convex and then concave upward as shown in D (**9,** pp. 317–319).

In addition to stimulus generalization, an analogous phenomenon of response generalization is sometimes supposed to operate so that (*a*) a stimulus to which one response has
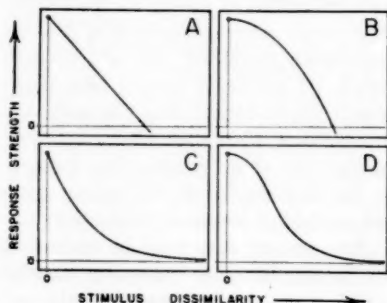
FIG. 1. Several proposals for the form of the gradient of stimulus generalization. The abscissa in each case is some measure of the dissimilarity between two stimuli, and the ordinate is some measure of the tendency of a response, conditioned to one of the two stimuli, to follow the presentation of the other.

242

been conditioned tends also to evoke other responses, and (b) the magnitude of this tendency (for any particular one of those responses) is governed by the dissimilarity between that response and the response originally conditioned. Although this principle of response generalization is also of considerable theoretical importance, even less progress has been made toward the quantitative determination of the shape of the gradient in this case than in the case of stimulus generalization.

Actually, that a unique gradient of generalization has failed to emerge both in studies of stimulus generalization and in studies of response generalization no longer seems so surprising when one examines these studies in detail. For although the obtained gradient must depend crucially upon the choice of the independent variable (i.e., dissimilarity), the various investigators have never been able to agree on any one measure of dissimilarity as appropriate. Furthermore, since tests for generalization have been conducted during various stages of discrimination learning, as well as after various periods of extinction, one must consider the possibility that the form of the gradient may change radically under different conditions of reinforcement (13, 31).

This paper will attempt to show that, by approaching the generalization problem from a somewhat different direction, considerable evidence can be adduced for the proposition that the gradients of stimulus and response generalization both conform to an exponential decay function (Curve C of Fig. 1). Further evidence will be presented to indicate that the form of the gradient does indeed depend upon the schedule of reinforcement and, more particularly, that it changes from the exponential to a bell-shaped

function (Curve D) as nonreinforced trials are introduced with greater and greater frequency.

## THE EXPONENTIAL GRADIENT IN PAIRED–ASSOCIATE LEARNING

In view of the difficulties attending any effort to establish one measure of dissimilarity as a standard, the following strategy was recently proposed: Stimuli (or responses) were conceptualized as points in a "psychological space" in such a way that the distance between any pair of these points represented the psychological dissimilarity between the corresponding pair of stimuli (or responses). By taking account of a set of metric axioms which any measure of distance should satisfy, it was shown that hypotheses about the shape of the gradient of generalization could be tested without resorting to an independent measure of dissimilarity (25). In a series of experiments on stimulus and response generalization during paired-associate learning, substantial support was obtained for the hypothesis that generalization is an exponential decay function of psychological distance (26). Since alternative hypotheses were not tested, however, it seems desirable to present data in a form that will clearly rule out the Functions A, B, and D of Fig. 1.

Now in a paired-associate experiment there are $N$ stimuli, $S_1, S_2, \ldots S_N$, to which are assigned $N$ responses, $R_1, R_2, \ldots R_N$. On each trial one of the $N$ stimuli is presented to the subject who must in turn produce one of the $N$ responses. Various procedures of differential reinforcement can be used to communicate to the subject the prevailing assignment of the responses to the stimuli. Sometimes the subject is simply informed after each trial whether his response was or

was not the correct one (i.e., the one assigned to the stimulus presented on that trial); at other times more elaborate methods of "correction" are used (26). In any case the so-called assignment is simply a rule which the experimenter follows in delivering the reinforcements, and any arbitrary one-to-one assignment can be established in this way.

Consider then a paired-associate experiment in which the responses are highly distinctive and so lead to negligible amounts of generalization. If there are $N$ pairs consisting of one stimulus and its assigned response, the data from such an experiment can be represented by an $N \times N$ matrix giving, for every $S_i$ and $S_k$, the conditional probability $P_{ik}{}^S$ with which $S_i$ leads to the response assigned to $S_k$ (25). A very basic measure of stimulus generalization between $S_i$ and $S_k$ is then provided by either of the probabilities $P_{ik}{}^S$ or $P_{ki}{}^S$. Comparison between different experiments is facilitated, however, by adjusting these measures so that the generalization between any stimulus and itself is always unity. This can be done by replacing the absolute probabilities $P_{ik}{}^S$ and $P_{ki}{}^S$ with the ratios $P_{ik}{}^S/P_{ii}{}^S$ and $P_{ki}{}^S/P_{kk}{}^S$. Furthermore, these two ratios can be averaged together to furnish a single, more stable estimate of the generalization between $S_i$ and $S_k$. Since there are theoretical reasons for preferring the geometric to the arithmetic mean (25), the measure of stimulus generalization might be defined by the formula

$$\sqrt{\frac{P_{ik}{}^S P_{ki}{}^S}{P_{ii}{}^S P_{kk}{}^S}}$$

Owing to the gradual manner in which differential reinforcement takes effect, however, there is an initial phase during which the subject re-

sponds more or less randomly. This means that the function given in the above formula necessarily levels off at some positive asymptote for large interstimulus distances. Again, in order to compare data from different experiments, this asymptote must be brought down to zero for each experiment. This is accomplished by estimating a parameter $C^S$ from each set of data (25), and by defining the generalization between $S_i$ and $S_k$ to be

$$G_{ik}{}^S = (1 + C^S)$$
$$\times \sqrt{\frac{P_{ik}{}^S P_{ki}{}^S}{P_{ii}{}^S P_{kk}{}^S}} - C^S \quad [1]$$

Likewise, in an experiment with highly discriminable stimuli, the response generalization between $R_i$ and $R_k$ will be given by

$$G_{ik}{}^R = (1 + C^R)$$
$$\times \sqrt{\frac{P_{ik}{}^R P_{ki}{}^R}{P_{ii}{}^R P_{kk}{}^R}} - C^R \quad [2]$$

where $P_{ik}{}^R$ is the conditional probability of $R_k$, given the stimulus assigned to $R_i$ (25).

Equations 1 and 2 then specify the dependent variables for paired-associate experiments, i.e., stimulus and response generalization. With regard to the independent variables, i.e., interstimulus and interresponse distances, it is clear that physical measures are not directly applicable. Thus the psychological distance between two tones at a fixed difference in intensity changes as both tones are increased in intensity. But this does not mean that there is no relation between physical and psychological distance. On the contrary, at least the following two statements can be made: (a) two tones which are brought arbitrarily close together in terms of physical measures also approach each other psychologically; (b) as two tones

separated by a fixed difference in intensity are increased in intensity, the psychological distance (although it does not remain fixed) nevertheless changes in a gradual and continuous manner. These considerations form a basis for the assumption that the locus of a set of stimuli in psychological space can always be obtained from their locus in physical space by a transformation that is (a) continuous and (b) differentiable (25).

Although the stimuli and responses chosen for experiments on paired-associate learning have typically been words or nonsense syllables, nonverbal materials can be used just as well. Indeed, for the purposes of establishing a measure of the psychological distance between stimuli, it is especially convenient to choose stimuli (such as tones differing only in intensity) that can be varied along a single physical dimension. In general, of course, stimuli that are evenly spaced along a single physical dimension will be neither evenly spaced nor confined to a straight line in psychological space. Nevertheless, it follows from the assumptions of continuity and differentiability that, if stimuli are evenly spaced over a sufficiently small range of a single physical dimension, then they will be spaced approximately evenly along an approximately straight line in psychological space.

Suppose, then, that the $N$ stimuli of a paired-associate experiment satisfy this special condition which, for present purposes, will be called the linearity condition. Such stimuli can be designated as $S_1, S_2, \ldots S_N$ in such a way that the subscripts correspond to the ordinal positions of the $N$ stimuli along the common (approximately straight) line. The average generalization for all pairs of stimuli just $D$ steps apart along this line will

then be given by

$$G(D) = \frac{1}{N - D} \sum G_{ik}$$

(with $i - k = D$) [3]

where, as indicated, the summation is carried out over the $N - D$ stimulus pairs separated by just $D$ steps. Similarly, if the responses meet the linearity condition, this formula can be used as a measure of the average generalization for all pairs of responses separated by $D$ steps. Thus, $G_{ik}$ in Formula 3 can stand for either $G_{ik}{}^S$ or $G_{ik}{}^R$.

Now the average generalization $G(D)$ and the separation $D$ have properties that particularly recommend them for investigations of the relation between generalization and distance. First of all, $D$ is quite insensitive to the physical measure used to space the stimuli evenly along the chosen dimension. Thus, if the stimuli are squares differing only in size, it does not much matter whether these squares are spaced in accordance with constant increments of area or length of side. It is a consequence of the linearity condition that these two variables ($a^2$ and $a$) will be almost linearly related within the small range of variation permitted along the size dimension. In fact, any variable that is equivalent to these variables except for a continuous, differentiable transformation could presumably be used. Thus the stimuli could just as well be regularly spaced in terms of measures based upon psychophysical procedures such as the summation of jnd's, category judgment, magnitude estimation, etc. (29). All of these measures are continuous and differentiable functions of the usual physical variables.

This indifference of $D$ to the variable chosen for evenly spacing the stimuli is further enhanced by com-
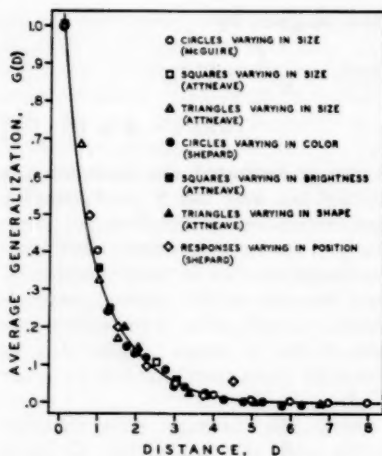
FIG. 2. An exponential decay function fitted to data from several experiments involving stimulus or response generalization during paired-associate learning. The center of each plotted symbol (triangle, square, circle) has average generalization $G(D)$ as ordinate and distance $D$ (multiplied by a constant factor) as abscissa. The experiments were conducted by Attneave (1), McGuire (18, 26), and Shepard (26). In the experiment with circles of variable color, the stimuli were not confined to a single physical dimension. Since the linearity condition was not met in this case, the $D$-values were taken from a multidimensional scaling solution obtained by Torgerson (30, Ch. 11). Since the stimuli were contained in a relatively small region of psychological space, it seems safe to assume that Torgerson's judgmental method of triads yields satisfactory estimates of psychological distance in the present sense. In the experiment on response generalization, it had been found that the two end-responses did not conform to the linearity requirement (26). The plotted data were therefore based upon the generalization between the seven intermediate responses only. Finally, Attneave's experiments deviated from the design presupposed by the present analysis in the following respects. (a) The stimuli were quite widely spaced and so may not have met the linearity requirement. (b) The number of trials was much smaller than in the other experiments. This makes the steady-state condition assumed in the ensuing theoretical discussion seem a little unlikely. (c) Since Attneave reported only the sums $P_{ik}{}^S + P_{ki}{}^S$ but not the individual

puting $G(D)$ as an average for all pairs of stimuli separated by $D$ steps. Thus, even if there is some residual systematic contraction or expansion of psychological distance, as one proceeds from the first to the last pair of stimuli at a given separation $D$, this systematic effect will be largely canceled out by averaging both ends of the stimulus range together.

In Fig. 2 average generalization $G(D)$ is plotted against $D$ for several experiments on the learning of paired associates. In each experiment several different random assignments of the responses to the stimuli were employed. Since the average psychological spacing of the stimuli and responses varies somewhat from experiment to experiment, the $D$-values for each experiment are multiplied by a constant factor to make them comparable with the $D$-values for the other experiments. The conclusion is clear: Although the empirical points fall closely along an exponential decay function (the fitted curve), they deviate markedly from the alternative Functions A, B, and D of Fig. 1.

## DEDUCTION OF THE EXPONENTIAL GRADIENT FROM A TRACE MODEL

The principal purpose of this paper is to propose a model to account for

probabilities, it was necessary to substitute the approximate formula $(P_{ik}{}^S + P_{ki}{}^S)/(P_{ii}{}^S + P_{kk}{}^S)$ for the geometric mean used in Equation 1. However, in a previous investigation this approximation was very close and, indeed, possessed greater statistical stability than the geometric mean (24). (d) There were not sufficient data to make the estimation of the constants $C^S$ feasible. For the purposes of plotting the data, therefore, $C^S$ was assumed to be zero. That the agreement between the various sets of data is good despite the deviations from optimum conditions in certain instances suggests that the data may not be particularly sensitive to such deviation.

the empirically determined exponential form of the gradient of generalization in terms of a hypothetical trace process. The aim of such a proposal is twofold: First, it is desired to remove the apparent arbitrariness of the exponential gradient by showing that it follows from certain elementary assumptions of a more intuitively compelling character. Second, it is hoped that such a model will provide for prediction to other experiments in which the usual paired-associate conditions no longer prevail.

The model is suggested by recognition experiments of the following kind. A subject is shown a certain square and, after a delay of $t$ units of time, shown a second square which may or may not be the same as the first. In this situation the probability of responding "same" is distributed, with respect to the difference in size between the first and second squares, according to some bell-shaped density function. Since the time error is usually small, the mode falls near the zero-difference point. The variance of the distribution, however, increases appreciably with the delay $t$ imposed between the first and second exposure (**2, 3, 10, 17**).

The unidimensional stimuli, $S_1$, $S_2$, . . . $S_N$, in an experiment on paired-associate learning, will be designated by small circles arranged in a vertical row along the left, as in A of Fig. 3. Corresponding to these, there are conceived the internal representations (perceptions), $S_1^*$, $S_2^*$, . . . $S_N^*$, which will be designated by the small circles on the right. Whenever a stimulus $S_i$ is presented, it leads to some one of the internal representations $S_k^*$, with probability $P_{ik}$. If the responses are so distinctive as never to be confused, reinforcement will ensure only if a stimulus is followed by its corresponding representation (**25**).
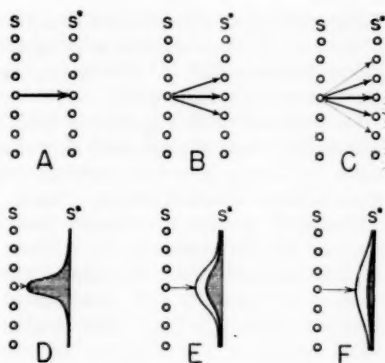


FIG. 3. The diffusion and deconditioning of the stimulus trace. The arrows connecting a single stimulus $S$ to the alternative perceptual representations $S^*$ in A, B, and C illustrate how the trace elements are assumed to spread out with time. In D, E, and F the process is represented as continuous.

Suppose, then, that when $S_i$ leads to $S_i^*$, a large number of trace elements are conditioned (by reinforcement) from $S_i$ to $S_i^*$. This bundle of trace elements or "stimulus trace" is designated by the arrow in A. Immediately following the removal of the external stimulus $S_i$, the trace elements are subject to haphazard perturbations so that some of the conditioned trace elements wander to adjacent stimulus representations as shown in B. (One might imagine here the action on the synapses of the random molecular processes associated with metabolism.) Later still, some of these elements will wander even further from $S_i^*$, as shown in C.

Now for any two stimuli differing along a single physical dimension, another stimulus can be found which is situated intermediately between these. Thus, as indicated in D, a continuum may be conceived as underlying the corresponding internal representations. This continuum then is the psychological space of the stimuli.

After removal of the stimulus, the distribution of trace elements progressively spreads out in this space, as illustrated in D, E, and F.

In addition to this spread or diffusion of the trace, it is assumed that the trace elements are also subject to spontaneous deconditioning (again, presumably owing to random processes at the molecular level). Thus, as the elapsed time increases, the number of elements still conditioned decays to zero. In Fig. 3, the shaded areas represent the fraction of the original trace elements which are still conditioned at each time.

### The Deconditioning of the Trace

The probability that a given conditioned element will suffer deconditioning during a small interval $\Delta t$ will be denoted by $U(\Delta t)$. The simplest rule that can be assumed to govern this probability is as follows:

*Assumption I.* $U(\Delta t)$ is a constant, independent of the time the element has remained conditioned and independent of the distance (in psychological space) to which the element has drifted in that time.

The probability that an element, still conditioned at time $t$, remains conditioned until $t + \Delta t$ is, of course, $1 - U(\Delta t)$. Therefore, the probability that a given trace element remains conditioned during the first $m$ intervals of length $\Delta t$, but then becomes deconditioned during the immediately succeeding $\Delta t$-interval, is, by Assumption I,

$$[1 - U(\Delta t)]^m U(\Delta t) \qquad [4]$$

Now $U(\Delta t)$ is necessarily proportional to the interval chosen for $\Delta t$. It is convenient, therefore, to define a deconditioning parameter,

$$U = \lim_{\Delta t \to 0} \frac{U(\Delta t)}{\Delta t} \qquad [5]$$

which does not depend upon the choice of an arbitrary interval $\Delta t$. Then, if the $t$-axis is translated so that conditioning takes place at $t = 0$, $m$ can be increased in such a way that, as $\Delta t \to 0$, $m \cdot \Delta t \to t$.

Although the probability that a given trace element is deconditioned at precisely time $t$ is zero, the probability per unit time (the "probability density") or rate of deconditioning for an individual trace element is, at the particular instant $t$,

$$\lim_{\Delta t \to 0} [1 - U(\Delta t)]^{t/\Delta t} U(\Delta t)/\Delta t$$

But by Equation 5 as $\Delta t \to 0$, $U(\Delta t) \to U \cdot \Delta t \to Ut/m$. Therefore, the rate of deconditioning at time $t$ is, for single trace elements,

$$\lim_{m \to \infty} [1 - Ut/m]^m U$$
$$= U \exp(-Ut) \qquad [6]$$

where, for convenience in what follows, $\exp(-Ut)$ is used in place of $e^{-Ut}$.

With regard to the deconditioned trace elements, the following rule is the simplest that can be assumed:

*Assumption II.* In the absence of further reinforcements, a deconditioned element remains deconditioned.

From this assumption and Equation 6 it follows that the fraction of the originally conditioned elements remaining conditioned at time $t$ is

$$U(t) = 1 - \int_0^t U \exp(-U\tau)d\tau$$
$$= \exp(-Ut) \qquad [7]$$

### The Diffusion of the Trace

Equation 7 completes the quantitative formulation of the deconditioning of the trace. In order to provide a similar formulation for the diffusion of the trace, it is necessary to consider

the motions of the individual trace elements in psychological space. The exposition is simplified by continuing to suppose that the stimuli are evenly spaced along a restricted range of a single physical dimension. It is then possible to introduce one coordinate $x_i$ for each stimulus representation $S_i$* giving its position along a one-dimensional psychological space. The psychological distance between any two stimuli, $S_i$ and $S_k$, is then

$$D_{ik} = |x_i - x_k| \qquad [8]$$

Now the expression $V_{ik}(\Delta x, \Delta t)$ will denote the probability that, if at time $t$ a trace element is situated at $x_i$, by time $t + \Delta t$ it will have moved into the one-dimensional region bounded by $x_k$ and $x_k + \Delta x$. As before, the arbitrary interval $\Delta x$ can be eliminated by defining a new quantity

$$V_{ik}(\Delta t) = \lim_{\Delta x \to 0} \frac{V_{ik}(\Delta x, \Delta t)}{\Delta x} \qquad [9]$$

$V_{ik}(\Delta t)$ is the probability density of displacements from $x_i$ to $x_k$ during the brief interval $\Delta t$. The simplest rule that can reasonably be assumed to govern this quantity is as follows:

*Assumption III.* $V_{ik}(\Delta t)$ is an invariant function of the psychological distance between $S_i$ and $S_k$. It does not depend upon the time $t$ or upon the absolute position of the pair, $S_i$ and $S_k$, in psychological space.

If the $x$-axis is translated so that $x_i = 0$, it follows from Assumption III that, for a given interval $\Delta t$, there exists a fixed function $f_{\Delta t}$, such that

$$V_{ik}(\Delta t) = f_{\Delta t}(D_{ik}) = f_{\Delta t}|x_k| \qquad [10]$$

However, it will not be necessary to make any particular assumption about the form of the function $f_{\Delta t}$. It is only necessary to insure that, during a short period of time, the probability of a very large displacement is negligible. This can be stated with greater precision as follows:

*Assumption IV.* The function $f_{\Delta t}$ and, thus, the probability density of displacements of a trace element from $x = 0$ is distributed over the $x$-axis with finite variance.

From Assumptions III and IV it follows that the variance of the distribution of displacements during an interval $\Delta t$ is the finite constant

$$V(\Delta t) = \int_{-\infty}^{+\infty} x_k^2 V_{ik}(\Delta t) dx_k \qquad [11]$$

Since $V(\Delta t)$ must depend upon the length of the interval $\Delta t$, it is useful to define a diffusion parameter,

$$V = \lim_{\Delta t \to 0} \frac{V(\Delta t)}{\Delta t} \qquad [12]$$

which does not require the stipulation of an interval $\Delta t$. Just as $U$ governs the rate of deconditioning of the trace, then, $V$ controls the rate of spread or diffusion of the trace in psychological space.

The next question to be answered is this: Given the diffusion parameter $V$, what form will the distribution of trace elements take after an appreciable delay $t$? It is possible to show that the assumptions which have been made are sufficient conditions for the desired distribution to tend toward a limiting form which is independent of the form of $f_{\Delta t}$ (**6, 15**). Specifically, if a trace comprising a large number ($n$) of elements is conditioned from $S_i$ to $x_i$ at $t = 0$, the density of these elements at $x_k$ for some later time $t$ conforms with the Gaussian function

$$n \cdot V_{ik}(t) = n \cdot (2\pi V t)^{-\frac{1}{2}}$$
$$\times \exp\left[-(D_{ik})^2/2Vt\right] \qquad [13]$$

The beauty of this result is its complete independence from the underlying mechanism symbolized by $f_{\Delta t}$. Thus, even though one imagines, for

example, that thermal agitation of the molecular substrate is responsible for deconditioning and diffusion, the biophysical details of this process need not be specified. For, according to the present formulation, these details are irrelevant to the question of the gross behavior of the trace system.

### The Trace Process in Paired-Associate Learning

Now in the course of learning paired associates, each stimulus $S_i$ will have been presented on many occasions. Furthermore, on a number of these occasions, reinforcement of the response assigned to $S_i$ will have conditioned a bundle of $n$ trace elements to $S_i^*$. Thus at some given time $t_0$, the density of conditioned elements resulting from the immediately preceding reinforcement will be distributed in psychological space as shown for $t_{-1}$ in Fig. 4. Likewise, the densities remaining from earlier reinforcements will be distributed as shown for $t_{-2}$, $t_{-3}$, and so on.

The total distribution of conditioned elements emanating from $S_i$ at $t_0$ can be found by summing the Gaussian distributions resulting from all previous reinforcements of the response assigned to $S_i$. It is possible to derive an analytic approximation to this composite distribution ($f_t$ in Fig. 4) by supposing that, after an

initial phase of rapid learning, reinforcements occur at relatively frequent and regular intervals. If the early trials are disregarded, then, a roughly steady-state process can be considered. The summation of the Gaussian curves of various ages can then be approximated by an integration of these curves over $t$.

Now, with respect to $S_i$, the density of trace elements at $x_k$ resulting from a reinforcement $t$ time units ago is given by $n \cdot V_{ik}(t)$. However, only the fraction $U(t)$ of these is still conditioned. Therefore the density of elements at $x_k$ which remain conditioned after a delay $t$ from reinforcement is $n \cdot U(t) \cdot V_{ik}(t)$. Clearly, then, the total density of conditioned trace elements resulting from all previous reinforcements of the response assigned to $S_i$ is distributed in approximate accord with

$$\int_0^\infty n \cdot U(t) \cdot V_{ik}(t) dt$$
$$= \int_0^\infty n \cdot \exp(-Ut) \cdot (2\pi Vt)^{-1}$$
$$\cdot \exp[-(D_{ik})^2/2Vt] dt \quad [14]$$

Fortunately, the integration can be effected (4, p. 144) and, indeed, yields

$$n(2UV)^{-\frac{1}{2}}$$
$$\cdot \exp[-D_{ik}(2U/V)^{\frac{1}{2}}] \quad [15]$$

For intermediate stages of paired-associate learning, then, this function can be taken as a measure of the strength of connection of the stimulus $S_i$ to the point $x_k$ on the perceptual continuum. It is not a probability density, however, since (if $n > 1$ or $U < 1$) the integration of this function over the $x$-axis yields a value greater than unity (7, p. 133). Multiplication of Equation 15 by $U/n$ converts this function to a prob-
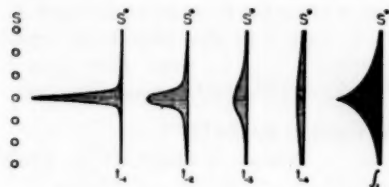


FIG. 4. A series of Gaussian distributions with increasing variances, and the composite distribution arising from an integration of these over time.

ability density,

$$P_{ik} = (U/2V)^{\frac{1}{2}}$$
$$\cdot \exp\left[- D_{ik}(2U/V)^{\frac{1}{2}}\right] \quad [16]$$

for then

$$\int_{-\infty}^{+\infty} P_{ik} dx_k = 1 \quad [17]$$

Now Equation 16 furnishes an estimate of the conditional probability (per unit $x$) of the particular perception $x_k$, given the stimulus $S_i$. In order to secure the probability of taking $S_i$ to be $S_k$ (through stimulus generalization), $S_k^*$ can be reinterpreted as a finite region partitioned off from psychological space in the neighborhood of $x_k$. In this way the entire one-dimensional space can be divided into $N$ mutually exclusive and exhaustive segments so that each segment corresponds to one of the external stimuli. A given trace element is then said to be conditioned to $S_i^*$ at time $t$ if and only if it falls in the region containing $x_i$ at that time.

$P_{ik}$, reinterpreted in this way, can be taken as an approximation to the conditional probability that the external stimulus $S_i$ will lead to the internal representation $S_k^*$. If the early trials are ignored (since these trials do not sufficiently approach a steady state), the constant $C^S$ in Equation 1 can be disregarded (25). Setting $C^S = 0$, and substituting the right-hand member of Equation 16 for $P_{ik}{}^S$ in Equation 1, then, the generalization between $S_i$ and $S_k$ assumes the remarkably simple form

$$G_{ik}{}^S = \exp\left(- D_{ik} \sqrt{2U/V}\right) \quad [18]$$

Letting $\kappa$ stand for the constant $\sqrt{2U/V}$ and averaging $G_{ik}{}^S$ over all pairs of stimuli separated by a fixed distance $D$, Equation 3 now takes the form

$$G(D) = e^{-\kappa D} \quad [19]$$

But if $\kappa$ is identified with the constant distance multiplier calculated for each experiment, this is precisely the exponential function fitted to the empirical data in Fig. 2.

*Further Aspects of the Trace Process*

*The role of deconditioning.* It might have seemed unnecessary to include the deconditioning assumptions (I and II) along with the diffusion assumptions (III and IV), since the process of diffusion alone would account for the spread of the trace, and hence for stimulus generalization. However, from Equation 18 it is clear that, without deconditioning, generalization would be so extensive as to prohibit the learning of paired associates. For if $U = 0$, $G_{ik} = 1$ for all $i$ and $k$. In this case the gradient is perfectly flat, and discrimination between stimuli is impossible. This is a consequence of carrying the integration of Equation 14 out to infinite $t$. Alternatively, one could integrate only out to some finite value $t = T$. However, in terms of the model, this is equivalent to assuming that the independent elements simultaneously suffer deconditioning at the same instant $T$ time units from conditioning. Rather than postulate coincidences of this kind, it seems more plausible to assume the gradual kind of fading away implied by Assumptions I and II. This fading away (or forgetting) then serves the adaptive function of weighting old, diffuse traces less heavily than new, accurate traces.

Of course, the integration of Equation 14 out to infinite $t$ is not strictly justified, since the learning experiment itself proceeds for only a finite period. The error introduced by the infinite integration is small, however, if the deconditioning parameter $U$ is sufficiently greater than zero. For

then the hypothetical traces persisting from reinforced trials that might be imagined as preceding the actual beginning of the experiment would be almost totally deconditioned during all but the early trials of the learning experiment.

*Asymmetric generalization.* It has sometimes been suggested that there exist asymmetries in which a generalization going in one direction, e.g., $S_i \rightarrow S_k{}^*$, is more probable than a generalization going in the reverse direction, $S_k \rightarrow S_i{}^*$. At first glance, such a possibility appears to violate Assumption III, according to which the probability of a displacement for a given trace element depends only upon the length (and not upon the direction) of that displacement. However, an account of such asymmetries which is consistent with Assumption III is suggested by the analysis proposed by Bush and Mosteller. According to their model, the psychological dissimilarity going from $S_i$ to $S_k$ is equal to the dissimilarity going from $S_k$ to $S_i$ only if the "set of stimulus elements" comprising $S_i$ has the same measure as the set of stimulus elements comprising $S_k$ (5). In terms of the trace model, then, it can be supposed that the area of the region of psychological space designated as $S_i{}^*$ may be greater or smaller than the area of the region designated as $S_k{}^*$. Since the probability that a given trace element (wandering at random) will occupy a certain region during an interval $\Delta t$ is proportional to the area of that region, $P_{ik}{}^S$ does not necessarily equal $P_{ki}{}^S$. Indeed, the weight $W_i{}^S$, proposed earlier by Shepard (25), is presumably a measure of the area of the region corresponding to $S_i$. This interpretation suggests why the empirically obtained weights tended to

be greatest for stimuli or responses at the end of a linear array (26).

*Multidimensional generalization.* For expository reasons the stimuli were supposed to vary along a restricted range of a single physical dimension. Actually, by making use of the theory of random motions in Euclidean spaces of more than one dimension (6), it can be shown that the trace model leads to the same exponential gradient in either of the two following multidimensional cases: (a) the psychological space is Euclidean; (b) the stimuli are confined to a small region of psychological space. (The second case follows from the first. For, by the hypothesized relation between physical and psychological space, a sufficiently small region of even a non-Euclidean space will be approximately Euclidean.) The treatment of generalization over large distances in non-Euclidean spaces, however, awaits the development of a general theory of random motions in spaces of positive and negative curvatures.

*Response generalization.* The preceding discussion has been formulated solely in terms of stimulus generalization. However, the same trace model can also be applied in the case of response generalization. Suppose, for example, that the responses are closely spaced along a single physical dimension, whereas the stimuli are completely discriminable. When one response $R_i$ is intended, some other response $R_k$ may actually be made. Using a notation analogous to that adopted for stimulus generalization, it can be said that $R_i{}^*$ leads to $R_k$ with probability $P_{ik}{}^R$. If then $R_i{}^*$ leads to $R_i$, the ensuing reinforcement conditions a large number of trace elements from $R_i{}^*$ to a point $x_i$ in the psychological space of the responses. These conditioned trace elements are

then subject to the rules already set forth in Assumptions I through IV.

## THE FORM OF THE GRADIENT AND THE SCHEDULE OF REINFORCEMENT

In the last section the apparent arbitrariness of the specific function fitted to the paired-associate data of Fig. 2 was reduced by showing that this function is a consequence of elementary assumptions which do not involve the specification of any particular function. The purpose of this section is to determine whether these same assumptions have any consequences for different types of experiments in which the sequence of reinforced and nonreinforced trials is manipulated, as in the study by Humphreys (13).

Certainly, from Fig. 4 it is clear that the sharp peak of the combined gradient at $x = 0$ is contributed solely by the distributions of trace elements resulting from the most recent reinforcements (like the distribution $t_{-1}$). Therefore, if feedback as to the correctness of each response is terminated, the composite gradient should become rounded (convex upward) in the vicinity of $x = 0$ and, under continued extinction, gradually flatten out. Likewise, if reinforcement is delivered only for very $n$th correct response, the composite gradient should assume an intermediate form which tends more toward either an exponential or a bell-shaped curve as $n$ is made smaller or larger. Considerations of this kind may account for the finding of Humphreys that, whereas with 100 per cent reinforcement the generalization gradient was concave upward, with 50 per cent reinforcement it was initially convex upward (13).

It is also feasible to analyze data from paired-associate experiments
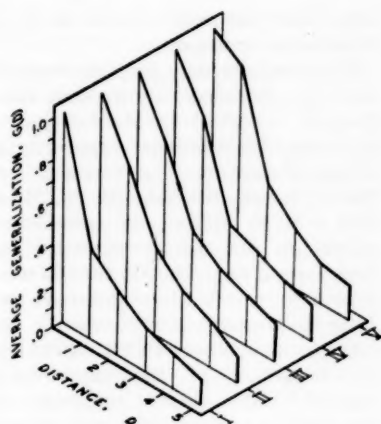


FIG. 5. The gradient of generalization as a function of the number of trials intervening between a response and the last preceding feedback as to the correctness of that response. The numbers of intervening trials are grouped together as follows: I. 0 trials; II. 1–3 trials; III. 4–8 trials; IV. 9–15 trials; and V. 16 or more trials.

with 100 per cent reinforcement for this effect. Since the stimuli are presented in random order, the delay between successive occurrences of a given stimulus or response varies over a considerable range. Thus separate gradients can be plotted depending on the number of trials intervening between a given stimulus-response sequence and the most recent feedback concerning the correctness of that particular sequence. The most extensive data currently available in a form suitable for this analysis come from an experiment on response generalization in paired-associate learning (26). Those data are therefore re-analyzed and plotted in Fig. 5. As predicted from the model, the gradient of generalization systematically changes from concave to convex upward in the vicinity of the correct response as the number of trials intervening between a response and the

last differential reinforcement of that response is increased.

These results may help to explain why the so-called "discriminal dispersion" observed in absolute-judgment and identification experiments seems to conform to a Gaussian or normal density function (**9**, pp. 317–319). Since differential reinforcements are not usually provided in these experiments, the discriminal dispersion is presumably maintained by those haphazard reinforcements of everyday existence which antedate the beginning of the experiment proper. Under these circumstances the gradient must be relatively mesokurtic, as illustrated in V of Fig. 5. This gradient resembles a Gaussian function, and is quite unlike the comparatively leptokurtic gradients (such as I) which have been shaped by recent differential reinforcement.

## MICROMECHANICAL AND MACROMECHANICAL MODELS FOR GENERALIZATION

The present trace model might be termed a *micromechanical model* because it is based upon assumptions about the fine-grain or "microscopic" mechanics of the underlying trace process. This model is to be distinguished from an earlier *macromechanical model* derived from assumptions pertaining to large-scale or "macroscopic" aspects of generalization (**25**). However, to the extent that the two models are compatible, they can be integrated so that the macromechanical assumptions of the earlier model will appear as deductions from the more primitive micromechanical assumptions of the present model. The aim of such an integration would be to broaden the scope of the earlier model. For, whereas the macromechanical model applies only to paired-associate experiments with continual differential reinforcement, the micromechanical model has implications for a considerably wider range of experiments.

The three basic assumptions from which the macromechanical model was originally derived are as follows. (*a*) Stimulus and response generalizations take place independently of each other. (*b*) The probability of a stimulus generalization is an exponential decay function of the psychological distance between the stimuli. (*c*) The probability of a response generalization is an exponential decay function of the psychological distance between the responses. Now to say that the scope of the earlier model will be increased by the adjunction of the micromechanical assumptions is to say that the macromechanical assumptions (*a*, *b*, and *c*) will retain their original form only under certain limiting conditions (e.g., under continual reinforcement). When these conditions are modified (as when the reinforcements are delivered only intermittently), Propositions *b* and *c* will have to assume different forms in accordance with the conclusions of the last section.

Even with continual reinforcement, there is one case of paired-associate learning for which the macromechanical assumptions would have a different form if deduced from the micromechanical assumptions. Specifically, if the interstimulus distances and the interresponse distances are both quite small, the trace model does not lead directly to the exponential gradient. Because occasional sequences of the form $S_i \rightarrow S_k{}^* \rightarrow R_k{}^* \rightarrow R_i$ will be reinforced, trace elements will be conditioned from $S_i$ to $x_k$ (rather than only to $x_i$). Events of this kind will somewhat alter the form of the gradient. (In a previous experiment with

generalization, both between the stimuli and between the responses, no significant departure from prediction on the assumption of an exponential gradient was observed [26]. However, the theoretically expected deviations would be quite small and may have been obscured by the rather large variability of the data from that experiment.)

Detailed derivations from the micromechanical assumptions are quite complicated in the case of simultaneous stimulus and response generalization, owing to the circumstance that, in this case, the form of the gradient depends upon the particular stimulus-response assignment enforced. Part of this complication is connected with the absence in both models of any account of the decrease in generalization which necessarily accompanies learning. An entirely satisfactory treatment of these problems will probably require an even more basic integration of these models with the already extensively developed models for learning per se.

## Summary

The problem of the relation between generalization and dissimilarity (i.e., the problem of the shape of the "gradient of generalization") is re-examined in the light of recent theoretical and empirical developments. With regard to experimental arrangements in which reinforcments are delivered in accordance with a one-to-one assignment of the responses to the stimuli (as in paired-associate learning), the following conclusions are drawn:

1. Measures of generalization can be defined in terms of the conditional probabilities with which the various stimuli lead to the various responses.

2. Thus defined, stimulus generalization and response generalization are both invariant functions of interstimulus and interresponse dissimilarities, respectively, provided that two conditions are met. First, dissimilarity is reinterpreted to mean a "psychological distance" which (a) is equivalent to "physical distance" except for a continuous, differentiable transformation, and (b) satisfies the metric axioms. Second, a given schedule of reinforcement is maintained.

3. Under conditions of frequent and regular reinforcement (as in the typical paired-associate experiment), the gradient of generalization is closely approximated by an exponential decay function (concave upward).

4. Under conditions of infrequent or intermittent reinforcement, this gradient departs from the exponential function in that it is convex upward in the vicinity of the reinforced stimulus or response.

5. The empirically observed gradients of generalization can be deduced from a mathematical model based upon four elementary assumptions concerning the temporal decay of stimulus and response traces.

## REFERENCES

1. Attneave, F. Dimensions of similarity. *Amer. J. Psychol.*, 1950, **63**, 516–556.
2. Bachem, A. Time factors in relative and absolute pitch determination. *J. acoust. Soc. Amer.*, 1954, **26**, 751–753.
3. Baldwin, J. M., & Shaw, W. J. Memory for square-size. *Psychol. Rev.*, 1895, **2**, 236–239.
4. Bierens de Haan, D. *Nouvelles tables d'intégrales définies.* New York: G. E. Stechert, 1939.
5. Bush, R. R., & Mosteller, F. A model for stimulus generalization and discrimination. *Psychol. Rev.*, 1951, **58**, 413–423.
6. Chandrasekhar, S. Stochastic problems in physics and astronomy. In N. Wax (Ed.), *Noise and stochastic processes.* New York: Dover, 1954. Pp. 3–89.

7. FELLER, W. *An introduction to probability theory and its applications.* New York: Wiley, 1950.

8. GRANDINE, LOIS, & HARLOW, H. F. Generalization of the characteristics of a single learned stimulus by monkeys. *J. comp. physiol. Psychol.*, 1948, **41**, 327–338.

9. GUILFORD, J. P. *Psychometric methods.* (2nd ed.) New York: McGraw-Hill, 1954.

10. HARRIS, J. D. The decline of pitch discrimination with time. *J. exp. Psychol.*, 1952, **43**, 96–99.

11. HOVLAND, C. I. The generalization of conditioned responses: I. The sensory generalization of conditioned responses with varying frequencies of tone. *J. gen. Psychol.*, 1937, **17**, 125–148.

12. HULL, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

13. HUMPHREYS, L. G. Generalization as a function of method of reinforcements. *J. exp. Psychol.*, 1939, **25**, 361–372.

14. JOHNSON, D. M. Generalization of a scale of values by the averaging of practice effects. *J. exp. Psychol.*, 1944, **34**, 425–436.

15. KAC, M. Random walk and the theory of Brownian motion. In N. Wax (Ed.), *Noise and stochastic processes.* New York: Dover, 1954. Pp. 369–391.

16. LASHLEY, K. S., & WADE, MARJORIE. The Pavlovian theory of generalization. *Psychol. Rev.*, 1946, **53**, 72–87.

17. LEYZOREK, M. Two-point discrimination in visual space as a function of the temporal interval between the stimuli. *J. exp. Psychol.*, 1951, **41**, 364–375.

18. McGUIRE, W. J. A multi-process model for paired-associates learning. Unpublished doctoral dissertation, Yale Univer., 1954.

19. MARGOLIUS, G. Stimulus generalization of an instrumental response as a function of the number of reinforced trials. *J. exp. Psychol.*, 1955, **49**, 105–111.

20. PLOTKIN, L. Stimulus generalization in Morse code learning. *Arch. Psychol.*, 1943, **40**, No. 287.

21. RAZRAN, G. Stimulus generalization of conditioned responses. *Psychol. Bull.*, 1949, **46**, 337–365.

22. ROTHKOPF, E. Z. A measure of stimulus similarity and errors in some paired-associate learning tasks. *J. exp. Psychol.*, 1957, **53**, 94–101.

23. SCHLOSBERG, H., & SOLOMON, R. L. Latency of response in a choice discrimination. *J. exp. Psychol.*, 1943, **33**, 22–39.

24. SHEPARD, R. N. Stimulus and response generalization during paired-associates learning. Unpublished doctoral dissertation, Yale Univer., 1955.

25. SHEPARD, R. N. Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika*, 1957, **22**, 325–345.

26. SHEPARD, R. N. Stimulus and response generalization: Tests of a model relating generalization to distance in psychological space. *J. exp. Psychol.*, 1958, **55**, 509–523.

27. SPENCE, K. W. The differential response in animals to stimuli varying within a single dimension. *Psychol. Rev.*, 1937, **44**, 430–444.

28. SPENCE, K. W. A reply to Dr. Razran on the transposition of response in discrimination experiments. *Psychol. Rev.*, 1939, **46**, 88–91.

29. STEVENS, S. S. On the psychophysical law. *Psychol. Rev.*, 1957, **64**, 153–181.

30. TORGERSON, W. S. *Theory and methods of scaling.* New York: Wiley, 1958.

31. WICKENS, D. D., SCHRODER, H. M., & SNIDE, J. D. Primary stimulus generalization of the GSR under two conditions. *J. exp. Psychol.*, 1954, **47**, 52–56.

---

---